Contents

0.1 Course Content

- Review of the basic principles of quantum mechanics
  - Foundations of quantum mechanics
  - Formalism
- Symmetries, esp. angular momentum
- Approximations
  - Variational Principle
  - Perturbation theory
- Identical particles, Second quantization
- Quantization of the electromagnetic fields, Interaction between light and matter
- Relativistic quantum mechanics
- Mixed states, entanglement, quantum information

0.2 Literature and text books

- *Quantum Mechanics*, Claude Cohen-Tannoudji, Wiley: Introduction into Quantum Mechanics
- *Quantum Physics*, Michel LeBellac, Cambridge University Press: From start of QM to some rather advanced concepts; very modern presentation with nice references to experimental verification
- *Modern Quantum Mechanics*, J.J. Sakurai, Addison Wesley
- *Avanced Quantum Mechanics*, J.J. Sakurai, Addison Wesley
- *Advanced Quantum Mechanics*, F. Schwabl, Springer
- *Grundkurs Theoretische Physik, Bd. 5*, W. Nolting, Springer (German-language standard texbook)
- *Fundamentals of Many-body Physics: Principles and Methods*, W. Nolting, Springer

0.3 Preliminary Schedule

This is a rough sketch of the topics assigned to their probably time in the course. We will certainly deviate from it. The empty slots mean that we will probably use that time for either the topic mentioned before or that mentioned after. [1]

Oct 16	Introduction
Oct 20	revision formalism: Hilbert spaces
Oct 23	time evolution, quantization
Oct 27	free non-relativistic particle: positive energy; harmonic oscillator
Oct 30	coherent states, symmetries, continuous translations
Nov 3	discrete transl. + tight-binding
Nov 6	rotations
Nov 9	finished symmetries, revision spin
Nov 13	Angular momentum, algebra

[1] They do not mean that there is no class on that day!

Nov 17	Angular momentum, algebra; scalars and vectors
Nov 20	Angular momentum addition
Nov 24	mixed states, density matrix, entanglement
Nov 27	EPR, variational principle, start time-indep. Perturbation theory
Dec 1	degenerate perturbation theory
Dec 4	Heisenberg/Dirac pictures
Dec 8	Time-dependent perturbation theory
Dec 11	quantized field theories: sound
Dec 15	quantization of electromagnetic field
Dec 18	Identical particles
Dec 22	"Second quantization", bosons
Jan 8	"Second quantization", fermions
Jan 12	singlet/triplet splitting for two electrons
Jan 15	Some condensed-matter stuff
Jan 19	Relativistic stuff: Mechanics, Klein-Gordon equation
Jan 22	Dirac equation
Jan 26	anti particles, non-relativistic limit, Pauli equation
Jan 29	Free particle
Feb 2	Symmetries in relativistic QM
Feb 9	

1 Revision

Maybe the Cohen-Tannoudji textbook

1.1 Formalism

1.1.1 States

Quantum mechanical systems are described in *Hilbert spaces*. The system's state is given by an element ("vector") of this complex vector space. Properties of a Hilbert space $\mathcal{H}$:

1. Linear: If ψ_1 and ψ_2 are elements of $\mathcal{H}$, then any $\varphi = \alpha\psi_1 + \beta\psi_2$ (for $\alpha, \beta \in \mathbb{C}$) are also in $\mathcal{H}$.
2. There is an "inner product": A mapping $c = \langle\varphi|\psi\rangle$ that maps a pair φ, ψ of vectors onto $\mathbb{C}$ and has the following properties:
   a) Swapping the elements of the pair gives the complex conjugate
   $$\langle\varphi|\psi\rangle = \langle\psi|\varphi\rangle^* \tag{1.1}$$
   b) Linear in the second element [1]
   $$\langle\varphi|\alpha\psi_1 + \beta\psi_2\rangle = \alpha\langle\varphi|\psi_1\rangle + \beta\langle\varphi|\psi_2\rangle \tag{1.2}$$
   c) The inner product of a vector with itself [which is $\in \mathbb{R}$, see Eq. (1.1)] is positiv definite:
   $$\langle\varphi|\varphi\rangle \geq 0 \qquad \text{and} \qquad \langle\varphi|\varphi\rangle = 0 \quad \Leftrightarrow |0\rangle = 0\ . \tag{1.3}$$
   The inner product is then used to define a norm $||\psi|| = \sqrt{\langle\psi|\psi\rangle}$ and a distance $d(x, y) = ||x - y||$
3. It is complete. This means that any Cauchy sequence [2] converges to a limit that is also contained in $\mathcal{H}$. (A counter example would be the rational numbers, where one can define sequences that converge towards irrational numbers. But one can fill these holes by adding the limits of sequences to the space.)
4. It is separable, meaning it has at most countably infinitely many dimensions. (A finite number of dimensions also works, of course.) Any element ϕ in the Hilbert space (in our case: any quantum state) can thus be expressed by a sum over a basis:
   $$\phi = \sum_m c_m \phi_m\ , \tag{1.4}$$

[1] At least in physics, in math, it is linear in the first element.

[2] This is a sequence of vectors, where one can choose an arbitrarily small cutoff and then finds that "almost all" (= all except a finite number of) elements are at most this far apart from each other.

where the ϕ_m are the basis vectors, which can be chosen to be normalized and orthogonal

$$\langle \phi_m | \phi_n \rangle = \delta_{m,n} \ . \tag{1.5}$$

Conversely, the coefficient c_m is obtained using the scalar product

$$\langle \phi_n | \phi \rangle = \sum_m c_m \underbrace{\langle \phi_n | \phi_m \rangle}_{=\delta_{n,m}} = c_n \ . \tag{1.6}$$

A useful example of a Hilbert space is the space of square-integrable functions, for which the following integral exists and is finite:

$$\int_x dx |f(x)|^2 = C < \infty \tag{1.7}$$

For the treatment of quantum mechanics, these functions offer the advantage of being normalizable.

1.1.1.1 Dirac notation

A commonly used notation, taken from the inner product, is to write the elements of the Hilbert space as "kets" $|\phi\rangle$. Each ket can also be used as the first element in the inner product, giving the corresponding "bra" $\langle \phi |$. (In the example of the square-integrable functions, the bra is the complex conjugate of the ket.) However, the "bra" notation can be extended to other linear functionals [3], which is used in extending the Hilbert space, see below.

1.1.2 Operators

Operators are mappings from one state to another, we are concerned with *linear* operators, i.e.:

$$A(|\phi\rangle + \lambda |\psi\rangle) = A|\phi\rangle + \lambda A|\psi\rangle \ . \tag{1.8}$$

Operators are not necessarily defined for the whole Hilbert space. An important related operator is the adjoint operator $A^\dagger$, which is defined by requiring that the following hold for all states for which the expression is defined:

$$\langle A\phi | \psi \rangle = \langle \phi | A^\dagger \psi \rangle \ . \tag{1.9}$$

Special operators are the zero- and one-operators:

$$0|\phi\rangle = 0 \cdot |\phi\rangle = 0 \tag{1.10}$$

$$\mathbb{I}|\phi\rangle = 1 \cdot |\phi\rangle = |\phi\rangle \ , \tag{1.11}$$

[3] A functional is a mapping from the vector space onto the complex numbers. Integrals are linear functional, the Dirac δ distribution is another example.

where a commonly used trick is to represent $\mathbb{I}$ as a sum over a basis, using (1.4) and (1.6):

$$\sum_n |n\rangle \underbrace{\langle n|\phi\rangle}_{\phi_n} = \sum_n \phi_n |n\rangle = |\phi\rangle \quad \Rightarrow \quad \sum_n |n\rangle\langle n| = \mathbb{I}\,. \tag{1.12}$$

Special kinds of operators are:

- For *self-adjoint* operators $A = A^\dagger$, including the requirement that they be defined on the same subspace. If said subspace is (or can be extended to) the full Hilbert space, A is *Hermitian*.
  - Their eigenvalues a_n are real. Let's assume that $|\phi_n\rangle$ is an eigenstate with eigenvalue a_n, i.e., $A|\phi_n\rangle = a_n|\phi_n\rangle$, then $A^\dagger = A$ implies:

$$\begin{aligned}\langle A\phi_n|\phi_n\rangle &= \langle \phi_n|A\phi_n\rangle^* = (a_n\langle\phi_n|\phi_n\rangle)^* = a_n^* \cdot 1 \\ &= \langle\phi_n|A^\dagger\phi_n\rangle = \langle\phi_n|A\phi_n\rangle = a_n \cdot 1, \text{ i.e. } a_n \in \mathcal{R}\end{aligned} \tag{1.13}$$

  - Eigenstates to different eigenvalues are orthogonal. Let's assume $A|\phi_n\rangle = a_n|\phi_n\rangle$ and $A|\phi_m\rangle = a_m|\phi_m\rangle$, then

$$\begin{aligned}\langle A\phi_n|\phi_m\rangle &= \langle\phi_m|A\phi_n\rangle^* = (a_n\langle\phi_m|\phi_n\rangle)^* = a_n^*\langle\phi_n|\phi_m\rangle \\ &= \langle\phi_n|A^\dagger\phi_m\rangle = \langle\phi_n|A\phi_m\rangle = a_m\langle\phi_n|\phi_m\rangle\,.\end{aligned} \tag{1.14}$$

    Either $a_n = a_m$ or $\langle\phi_n|\phi_m\rangle = 0$ has to hold.

  Their eigenstates actually form a basis and one can thus use this basis to express quantum states, if convenient. Physical observables are Hermitian operators $A = A^\dagger$.
- Another important class of operators are ones that conserve the scalar product (resp. the norm): $\langle U\phi|U\psi\rangle = \langle\phi|\psi\rangle$. For *unitary operators*, $U^{-1} = U^\dagger$ has to hold. Concerning eigenvalues, they tend to be as well behaved as Hermitian operators, except that eigenvalues can be complex, but have to be of unit length, i.e., of the form $u_n = \mathrm{e}^{i\phi_n}$ with real ϕ_n. Unitary operators are often used to express symmetries of the system.

If and only if two observable commute, i.e., if $[A, B] = AB - BA = 0$, an eigensystem common to both can be found. Let us look at the two directions of this statement:

- Let's assume that A and B have a common eigensystem $|\psi_n\rangle$, with eigenvalues a_n and b_n, respectively. Any state $|\phi\rangle$ can be expressed in this basis and the commutator $[A, B] = AB - BA$ becomes

$$\begin{aligned}[A,B]|\psi\rangle = (AB-BA)\sum_n c_n|\psi_n\rangle = \sum_n c_n(AB-BA)|\psi_n\rangle = \sum_n c_n(Ab_n|\psi_n\rangle - Ba_n|\psi_n\rangle) = \\ = \sum_n c_n(b_nA|\psi_n\rangle - a_nB|\psi_n\rangle) = \sum_n c_n(b_na_n - a_nb_n)|\psi_n\rangle = 0\,,\end{aligned} \tag{1.15}$$

  where we used that A and B are linear operators so that they always commute with a number.

- Let us now assume that $[A, B] = 0$ as an operator identity, i.e., when applied to an arbitrary state. Thus $AB = BA$, which we apply to an eigenstate $|a_n\rangle$ of A:

$$AB|a_n\rangle = BA|a_n\rangle = Ba_n|a_n\rangle = a_n|Ba_n\rangle \tag{1.16}$$

  The state $B|a_n\rangle$ is thus likewise an eigenstate of A, with the same eigenvalue a_n. There are now two possibilities:
  - Only one eigenstate has eigenvalue a_n, i.e., a_n is *non degenerate*. $B|a_n\rangle$ then has to be "essentially" the same state as $|a_n\rangle$, i.e., the two have to be linearly dependent, meaning $B|a_n\rangle = b_n|a_n\rangle$. This means that $|a_n\rangle$ is in fact an eigenstate of B as well.
  - a_n is *degenerate*, i.e., a subspace of the Hilbert space has this eigenvalue. The consideration above implies that applying B to a state in this subspace will lead to another vector of the subspace. As any basis of this subspace is an eigenbasis of A, we just choose one that coincides with the eigenbasis of B in this subspace.
- Note that we have not used that A and B are Hermitian, nor that the eigenvalues are real. The statement about common eigensystems and commuting operators can thus be extended to a Hermitian operator (e.g. the Hamiltonian H) commuting with a unitary one (e.g, on expressing a symmetry), so that symmetries can be used to (partially) diagonalize a Hamiltonian.

1.1.3 States and operators as vectors and matrices

In finite-dimensional Hilbert spaces, states are vectors, see (1.4) and linear operators can be represented as matrices. This picture of matrices and vectors mostly carries over to Hilbert spaces with infinite dimensions, even though some math is needed for this. The matrix element can be obtained by 'sandwiching' the operator. We start from Eq. (1.4), apply the operator A and insert a $\mathbb{I}$ twice:

$$\begin{aligned} A|\phi\rangle = \mathbb{I}A\mathbb{I}|\phi\rangle = \left(\sum_n |n\rangle\langle n|\right) A \left(\sum_{n'} |n'\rangle\langle n'|\right) \left(\sum_m \phi_m |m\rangle\right) = \\ = \sum_{n,n',m} |n\rangle \underbrace{\langle n|A|n'\rangle}_{A_{n,n'}} \phi_m \underbrace{\langle n'|m\rangle}_{=\delta_{n',m}} = \sum_n |n\rangle \sum_m A_{n,m}\phi_m \end{aligned} \tag{1.17}$$

State $|\phi\rangle$ is thus represented by the vector with coefficients ϕ_m and operator A by the matrix with elements $A_{n,m}$ and applying A to $|\phi\rangle$ amounts to a matrix-vector multiplication. (Analogously, the product of two linear operators is given by a matrix product.)

1.1.4 Generalization

Nolting uses a the large-but-finite volume trick, LeBellac nicely discusses these aspects.

In fact, the situation in quantum mechanics is more tricky than that: Hilbert spaces with a countable number of dimensions a not enough to describe, e.g., real positions in space, because the real numbers are not countable. This can be fixed by including appropriate "generalized" states, the full thing is called a rigged Hilbert space. If one deals with a basis of these states, e.g., the eigenbasis of position and momentum operators, the usual sums are

replaced by integrals and orthonormality defined via a Dirac delta-distribution rather than a Kronecker delta.

Examples are functions like $\psi_k(x) = \mathrm{e}^{ikx}$ for $x \in \mathbb{R}$, which are the eigenfunctions of the momentum operator. They are not square integrable, but the scalar product $\langle\psi_k|\phi\rangle$ with a square integrable function is defined and finite. Moreover, all square integrable functions can be expressed in terms of the the ψ_k, as

$$\phi(x) = \int_k dk\; c(k)\psi_k(x) \tag{1.18}$$

with suitably chosen "coefficients" c_k. Even though these functions cannot be normalized if we assume an infinite volume, one can make a consistent orthonormal basis out of them:

$$\langle\psi_k|\psi_{k'}\rangle = \delta(k - k') \tag{1.19}$$

Similarly, the eigenfunctions of the position operator are not even proper functions, but are given by Dirac delta distributions $\delta(x)$.

1.2 Schrödinger Equation

1.2.1 Dynamics

For non-relativistic particle, time evolution of a state $|\psi\rangle$ obeys The Schrödinger equation

$$i\hbar\frac{d\,|\psi\rangle}{dt} = \hat{H}|\psi\rangle\ , \tag{1.20}$$

where $\hat{H}$ is the Hamilton operator. For a Hamiltonian that does not change with time $\partial\hat{H}/\partial t = 0$, this is formally easily integrated to

$$|\psi(t)\rangle = \mathrm{e}^{-\frac{i}{\hbar}\hat{H}t}|\psi(t=0)\rangle\ . \tag{1.21}$$

The exponential of the operator $\hat{H}$ is best expressed in terms of its eigenstates $|n\rangle$ with eigenvalues E_n, leading to

$$|\psi(t)\rangle = \sum_n \mathrm{e}^{-\frac{iE_n}{\hbar}t}|n\rangle\langle n|\psi(t=0)\rangle \tag{1.22}$$

for a Hamiltonian with a discrete spectrum, i.e., with a countable number of eigenstates. In the case of a continuous spectrum, we would get an integral

$$|\psi(t)\rangle = \int_p \mathrm{d}p\; \mathrm{e}^{-\frac{iE_p}{\hbar}t}|p\rangle\langle p|\psi(t=0)\rangle\ . \tag{1.23}$$

The states $|p\rangle$ are examples of the "generalized" states mentioned above. Note that the spectrum of a Hamiltonian can have both discrete and continuous parts, in this case, its full spectral decomposition includes both a the respective sum and integral.

We are consequently very interested in finding the eigenstates and eigenvalues of the Hamilton operator, i.e., in finding the solutions of the eigenvalue equation

$$\hat{H}|\psi\rangle = E|\psi\rangle\ . \tag{1.24}$$

A quantum-mechanical system that is in an eigenstate of H is in a 'stationary' state, because it remains in this state for all times and only picks up a phase. Alternatively, a system can initially be in a superposition of more than one eigenstate. During its time evolution, it can evolve into a state that is even orthogonal to the initial state.

1.2.2 Time-evolution operator

It is well known that the time-dependence of a state is described by the time-dependent Schrödinger equation, see Sec. 1.2. For a Hamiltonians that do not explicitly depend on time $\partial_t H = 0$, time evolution is given by Eq. (1.22). As one can easily verify, the corresponding operator $U(t, t_0 = 0)$ is unitary

$$|\psi(t)\rangle = \sum_n \mathrm{e}^{-\frac{iE_n}{\hbar}t}|n\rangle\langle n|\psi(t=0)\rangle = U(t, t_0 = 0)|\psi(t=0)\rangle \;, \tag{1.25}$$

i.e., $U^{-1}(t, t_0 = 0) = U^\dagger(t, t_0 = 0)$. (Moreover, $U^{-1}(t, t_0) = U(t_0, t)$ and $U(t, t_0) = U(t - t_0)$ only depends on the time *difference*.)

Even for time-dependent Hamiltonians, the time-evolution operator is unitary. This can be seen by noting that arbitrary scalar products are not changed by time-evolution:

$$\frac{d}{dt}\langle\phi(t)|\psi(t)\rangle = \left(\frac{d}{dt}\langle\phi(t)|\right)\psi(t)\rangle + \langle\phi(t)|\left(\frac{d}{dt}\psi(t)\rangle\right) \overset{(1.22)}{=} \frac{i}{\hbar}\langle\phi(t)|H|\psi(t)\rangle - \frac{i}{\hbar}\langle\phi(t)|H|\psi(t)\rangle = 0 \;. \tag{1.26}$$

As the time-evolution operator describing this time evolution preserves scalar products, it fulfils the definition of a unitary operator.

Let us now look at the time dependence of this operator itself. Without a time difference, the operator must become unity $U(t, t) = 1$. Moreover, one can decompose the time evolution into parts, i.e., $U(t, t_0) = U(t, t_1)U(t_1, t_0)$. Finally, as the wave function has a smooth time evolution, it makes sense to define

$$\partial_t U(t, t_0) = \lim_{dt\to 0} \frac{U(t + dt, t_0) - U(t, t_0)}{dt} = \lim_{dt\to 0} \frac{U(t + dt, t) - 1}{dt} U(t, t_0) \;. \tag{1.27}$$

We now look at the wave function again

$$i\hbar|\dot\psi(t)\rangle = i\hbar\partial_t|\psi(t)\rangle = i\hbar\partial_t U(t, t_0) \underbrace{|\psi(t_0)\rangle}_{\text{not } t\text{-dep.}} \tag{1.28}$$

and comparison to the Schrödinger equation yields

$$\hat H|\psi(t)\rangle = \hat H U(t, t_0)|\psi(t_0)\rangle = i\hbar\partial_t U(t, t_0)|\psi(t_0)\rangle \;. \tag{1.29}$$

As the Schrödinger equation works for all wave functions $|\psi(t_0)\rangle$ (for which $\hat H$ is defined), we can write it as an operator identity

$$i\hbar\partial_t U(t, t_0) = \hat H U(t, t_0) \;. \tag{1.30}$$

This is the time evolution of the unitary time-evolution operator. We have not used $\partial_t \hat H = 0$ here and this equation also holds for time-dependent Hamiltonians, even though its integration becomes more difficult.

A note on ∂_t vs. dt: The symbol ∂_t for the time derivative in the Schrdinger euqation (1.25) arose, because the state written as a wave function is a function of other variables (e.g. position) in addition to time t. However, the Schrödinger equation describes the full time dependence of the state and $d|\psi(t)\rangle/dt$ or $|\dot{\psi}(t)\rangle$ are consequently also valid notations. Similarly, we have here also included the full dependence of $U(t, t_0)$ on its first argument t. The notation $\partial_t U$ may be useful to clarify that it is the first time argument and not the second that is used in the differentiation. If no confusion is possible on this account, $d_t U$ can be used. In the case of the Hamiltonian, the notation $\partial_t H$ instead of dH/dt will, however, become meaningful in Sec. 6, where we will discuss the "Heisenberg picture". Operators can be seen to have an "implicit" time dependence that is due to the time-evolution and an "explicit" one that is imposed from outside. The symbol ∂_t will then refer to "explicit" time dependence.

1.2.3 Quantization

One way to obtain the Hamilton operator of a given system is via "quantization" of its classical Hamiltonian using the canonical position and momentum variables and replacing them by operators. Position and momentum operators $\hat{X}$ and $\hat{P}$ obey the canonical commutation relation

$$[\hat{P}, \hat{X}] = -i\hbar\hat{\mathbb{I}} \,. \tag{1.31}$$

In position space, they are given by

$$\hat{X}|\psi\rangle = \vec{r}|\psi\rangle \qquad \text{and} \tag{1.32}$$

$$\hat{P}|\psi\rangle = -i\hbar\nabla|\psi\rangle \,, \tag{1.33}$$

where $\vec{r}$ is the vector (x, y, z) denoting position and ∇ the gradient $(\partial_x, \partial_y, \partial_z)$ [4].

1.3 Free Particle (in one Dimension)

A "free" particle is one not feeling any forces, i.e., its classical energy and Hamiltonian are just given by its kinetic energy $E = \frac{\vec{p}^2}{2m}$. Using the rules (1.32) from above and selecting to work in position space [i.e. $|\psi(t)\rangle \to \psi(t, x)$], the Schrödinger equation becomes

$$i\hbar\frac{d\psi(t,x)}{dt} = \hat{H}\psi(t,x) = \frac{1}{2m}(-i\hbar\nabla)^2\psi(t,x) = -\frac{\hbar^2}{2m}\Delta\psi(t,x) \,, \tag{1.34}$$

with $\Delta = \partial_x^2 + \partial_y^2 + \partial_z^2$. $\psi(x)$ is here supposed to indicate that the state is written as a function of position x, often, a notation $\psi(x)$ is chosen whenever a basis has been committed to, in "bra"-"ket" notation, $\psi(x) = \langle x|\psi\rangle$.

We are now interested in solutions to the time-independent Schrödinger equation (1.24), and will for simplicity consider only one spatial dimension here.[5] We are looking for solutions

[4] In fact, this is one point where the 'tricky issues' with going beyond finite dimensions does play a role: $\hat{X}$ and $\hat{P}$ would both be traceless matrices and in finite dimensions, their commutator can then not have non-zero trace, i.e., they would not be able to fulfill (1.31).

[5] The solution in three dimensions is equivalent, as the equation can be factored into three equivalent equations, one for each dimensions.

to

$$-\frac{\hbar^2}{2m}\partial_x^2\psi(x) = E\,\psi(x) \quad \text{resp.} \quad \partial_x^2\psi(x) = -\frac{2mE}{\hbar^2}\,\psi(x)\,. \tag{1.35}$$

For each energy $E > 0$, there are two linearly independent solutions

$$\langle x|\psi_{E,+}\rangle = \psi_{E,+}(x) = \mathrm{e}^{i\frac{\sqrt{2mE}}{\hbar}x} \quad \text{and} \quad \langle x|\psi_{E,-}\rangle = \psi_{E,-}(x) = \mathrm{e}^{-i\frac{\sqrt{2mE}}{\hbar}x}\,, \tag{1.36}$$

the spectrum is thus continuous and doubly degenerate. The eigenfunctions cannot be normalized and there are more than countably many, we have here to deal with the "generalized" situation of Sec. 1.1.4 again. The time-dependent wave functions are

$$\psi_{E,+}(t,x) = \mathrm{e}^{\frac{i}{\hbar}(-Et+\sqrt{2mE}x)} \quad \text{and} \quad \psi_{E,-}(t,x) = \mathrm{e}^{\frac{i}{\hbar}(-Et-\sqrt{2mE}x)}\,. \tag{1.37}$$

1.3.1 Free particle in momentum space

While working in position space maybe the first thing to think about, the free particle can at least as easily be treated in momentum space. The operator $\hat{P}$ is here just multiplication with the variable p (or the vector $\vec{p}$ in higher dimensions). The eigenvalue equation in momentum space is then simply

$$\frac{p^2}{2m}\psi(p) = E\psi(p), \quad \text{resp.} \tag{1.38}$$

$$(p^2 - 2mE)\psi(p) = \left(p - \sqrt{2mE}\right)\left(p + \sqrt{2mE}\right)\,\psi(p) = 0\,. \tag{1.39}$$

The solutions are again a bit special, as they are not "normal" functions of p, but Dirac delta-distributions $\langle p|\psi_{E,\pm}\rangle = \psi_{E,\pm}(p) = A_\pm\delta(p \mp \sqrt{2mE})$. The momentum variable p is real, as it has to be given that it is the eigenvalue of the Hermitian momentum operator; this precludes any solutions $\neq 0$ for $E < 0$. Again, we find two solution per energy, one with positive and one with negative momentum p.

We could now go from momentum to position space, this is done by inserting a $\mathbb{I}$-operator in the form of a sum over the momentum eigenstates. As these are generalized states, we need an integral:

$$\psi_{E,\pm}(x) = A_\pm\langle x|\psi_{E,\pm}\rangle = A_\pm\langle x|\underbrace{\int_p \mathrm{d}p\;|p\rangle\langle p|}_{=\mathbb{I}}|\psi_{E,\pm}\rangle = A_\pm\int_p \mathrm{d}p\;\underbrace{\langle x|p\rangle}_{\frac{1}{\sqrt{2\pi\hbar}}\mathrm{e}^{i\frac{p}{\hbar}x}}\underbrace{\langle p|\psi_{E,\pm}\rangle}_{\psi_{E,\pm}(p)=\delta(p\mp\sqrt{2mE})} =$$

$$= \frac{A_\pm}{\sqrt{2\pi\hbar}}\int_p \mathrm{d}p\;\mathrm{e}^{i\frac{p}{\hbar}x}\delta(p \mp \sqrt{2mE}) = \frac{A_\pm}{\sqrt{2\pi\hbar}}\mathrm{e}^{\pm i\frac{\sqrt{2mE}}{\hbar}x}\,. \tag{1.40}$$

Fortunately, the solutions are the same as the ones we obtained directly in position space (1.41). Setting their normalization $A_\pm$ to 1 yields normalized (generalized) eigenstates with $\langle\psi_{E,\pm}|\psi_{E',\pm}\rangle = \delta(E - E')$. The way to go from momentum to position space (and back) is naturally the (inverse) Fourier transform.

1.3.2 Allowed energies

Allowed energies for the free particle are *positive*. When coming from momentum space, this is immediately clear, because it is given as a function of the momentum eigenvalue p as $E = \frac{p^2}{2m}$. Since p is the eigenvalue of a Hermitian operator, it is real and its square is positive.

When coming from position space, two solutions for $E < 0$, the two would be

$$\psi_{E,+}(x) = \mathrm{e}^{\frac{\sqrt{2mE}}{\hbar}x} \quad \text{and} \quad \psi_{E,-}(x) = \mathrm{e}^{-\frac{\sqrt{2mE}}{\hbar}x} \,. \tag{1.41}$$

Not only can they not be normalized, but they actually diverge strongly for $x \to \infty$ or for $x \to -\infty$. Consequently, they can not be used to define a valid bra, i.e., one cannot even get a valid generalized state from them. (If one uses the trick of a large by finite volume instead, these solutions are a problem because they can not be made periodic.)

1.4 Piecewise Constant potentials in one Dimension

see exercises

1.5 The Harmonic Oscillator in one Dimension

The Hamiltonian of the classical harmonic oscillator describes a particle moving in a square well and is given by kinetic and potential energy

$$T = \frac{m\dot{q}^2}{2}, \quad V = \frac{V_0}{2}q^2 = \frac{m\omega^2}{2}q^2 \,, \tag{1.42}$$

which give the Lagrangian

$$\mathcal{L} = T - V = \frac{m\dot{q}^2}{2} - \frac{m\omega^2}{2}q^2 \,. \tag{1.43}$$

A change of variables to $x = \sqrt{\frac{m\omega}{\hbar}}q$ gives

$$\mathcal{L} = \frac{\hbar\dot{x}^2}{2\omega} - \frac{\hbar\omega}{2}x^2 \,. \tag{1.44}$$

We identify here the canonical momentum corresponding to the new variable x:

$$p = \frac{\partial\mathcal{L}}{\partial\dot{x}} = \frac{\hbar\dot{x}}{\omega} \quad \Rightarrow \quad \dot{x} = \frac{\omega}{\hbar}p \,. \tag{1.45}$$

We then obtain the classical Hamiltonian in its canonical variables x and p as

$$H(x,p) = p\dot{x} - \mathcal{L} = \frac{\omega}{\hbar}p^2 - \frac{\hbar\dot{x}^2}{2\omega} + \frac{\hbar\omega}{2}x^2 = \frac{\hbar\omega}{2}\left(\frac{p^2}{\hbar^2} + x^2\right) \tag{1.46}$$

1.5.1 Algebraic analysis of the eigenvalues

The quantum-mechanical Hamiltonian can then be obtained by the "canonical quantization", where the canonical variables x and p are replaced by operators fulfilling the canonical commutation relation

$$[p, x] = -i\hbar\mathbb{I} \,. \tag{1.47}$$

One way to proceed would then be to solve the eigenvalue equation in position space, where $p = -i\hbar\partial_x$, and

$$\frac{\hbar\omega}{2}\left(-\partial_x^2 + x^2\right)\psi(x) = E\psi(x) \,. \tag{1.48}$$

A different way is the "algebraic" method which introduces operators

$$a = \frac{1}{\sqrt{2}}(x + \frac{i}{\hbar}p) \quad \text{and} \quad a^\dagger = \frac{1}{\sqrt{2}}(x - \frac{i}{\hbar}p) \,. \tag{1.49}$$

If x and p were *numbers*, then the Hamiltonian would be given by the product of these operators. As they are in fact operators, this does not quite work, but it still makes sense to look at the product:

$$a^\dagger a = \frac{1}{2}(x^2 + \frac{i}{\hbar}\frac{-i}{\hbar}p^2 - \frac{i}{\hbar}\underbrace{(px - xp)}_{=[p,x]=-i\hbar}) = \frac{1}{2}\left(x^2 + \frac{p^2}{\hbar^2} - 1\right) \tag{1.50}$$

and the rewrite the Hamiltonian as

$$H = \hbar\omega\left(a^\dagger a + \frac{1}{2}\right) = \hbar\omega\left(\hat{N} + \frac{1}{2}\right) \,. \tag{1.51}$$

The – so far unknown – eigenstates $|n\rangle$ of $\hat{N}$ and H are of course the same, with eigenvalues n and $\hbar\omega(n + 1/2)$. $\hat{N}$ is so far just a name for the combined operator $a^\dagger a$. Classically (i.e. for numbers x and p), $aa^\dagger$ would of course give the same. Consequently, we also look at the commutator:

$$[a, a^\dagger] = \frac{1}{2}\left(x^2 + \frac{p^2}{\hbar^2} + \frac{i}{\hbar}[p, x]\right) - \frac{1}{2}\left(x^2 + \frac{p^2}{\hbar^2} - \frac{i}{\hbar}[p, x]\right) = \frac{1}{2}\left(2\frac{i}{\hbar}[p, x]\right) = 1 \tag{1.52}$$

While we do not yet know the eigenvectors and -values, we can learn something about them by seeing how $a^\dagger$ and a act on them.

$$\begin{aligned} n \cdot a^\dagger|n\rangle = a^\dagger n|n\rangle = a^\dagger\hat{N}|n\rangle = a^\dagger \underbrace{a^\dagger a}_{aa^\dagger - 1}|n\rangle = (a^\dagger a - 1)a^\dagger|n\rangle \\ \Rightarrow \quad (n+1)a^\dagger|n\rangle = \hat{N}a^\dagger|n\rangle \end{aligned} \tag{1.53}$$

If $|n\rangle$ is an eigenvector corresponding to eigenvalue n, then $a^\dagger|n\rangle$ is consequently also an eigenvector and corresponds to eigenvalue $n + 1$. Analogously, we find that $a|n\rangle$ leads to an eigenvector with reduced eigenvalue:

$$n \cdot a|n\rangle = an|n\rangle = a\hat{N}|n\rangle = \underbrace{aa^\dagger}_{a^\dagger a + 1} a|n\rangle = (a^\dagger a + 1)a|n\rangle$$

$$\Rightarrow \quad (n-1)a|n\rangle = \hat{N}a|n\rangle \; . \tag{1.54}$$

The operator $a^\dagger$ (raising operator) and a (lowering operator) applied to an eigenstate lead to other eigenstates with higher or lower eigenvalue. In order to compute the length of the new vectors, we start by assuming that $|n\rangle$ has length 1, $||n|| = \sqrt{\langle n|n\rangle} = 1$.

$$||a|n\rangle||^2 = \langle an|an\rangle = \langle n|\underbrace{a^\dagger a}_{\hat{N}}|n\rangle = \langle n|n|n\rangle = n\underbrace{\langle n|n\rangle}_{1} \tag{1.55}$$

The norm of $a|n\rangle$ is thus $\sqrt{n}$, while "the" eigenvector $|n-1\rangle$ is the normalized one with length 1, i.e.,

$$|n-1\rangle = \frac{1}{\sqrt{n}}a|n\rangle, \quad \text{and} \quad |n+1\rangle = \frac{1}{\sqrt{n+1}}a^\dagger|n\rangle \; . \tag{1.56}$$

Obviously, one can apply the "ladder operators" several times, e.g., $|n-2\rangle = \frac{1}{\sqrt{n-1}}\frac{1}{\sqrt{n}}a^2|n\rangle$, which takes us from eigenvalue n to eigenvalue $n-2$. However, it turns out that the eigenvalues cannot be negative, because they can be written as the norm of a vector:

$$n = \langle n|\hat{N}|n\rangle = \langle n|a^\dagger a|n\rangle = ||a|n\rangle||^2 \geq 0 \tag{1.57}$$

It must thus not be possible to go to eigenvalues that are too small. The solution can be found by looking at

$$a^k|n\rangle = \sqrt{n(n-1)(n-2)\dots(n-k+1)}|n-k\rangle \tag{1.58}$$

This vector is proportional to the eigenvector for eigenvalue $n-k$. For $n-k<0$, the only way to satisfy Eq. (1.57) is for this vector to vanish, implying the nonexistence of the eigenvalue. Fir *integer* n, this works out, as the prefactor $\sqrt{\dots}$ becomes 0 for $k \geq n+1$. For non-integer n, however, we would find a contradiction: The prefactor would not contain 0 for $k > n$, vector (1.58) would have finite length and the negative eigenvalue $n-k$ would thus exist, in contradiction to Eq. (1.57). *All eigenvalues of the number operator $\hat{N} = a^\dagger a$ are thus integer and non-negative.* Eigenenergies of the harmonic oscillator are given by

$$\hat{H}|n\rangle = \hbar\omega\left(n+\frac{1}{2}\right)|n\rangle \tag{1.59}$$

While there is a smallest eigenvalue $n = 0$, there is no largest one: applying $a^\dagger$ repeatedly always leads to eigenvectors with higher eigenvalue, and the prefactor $\sqrt{(n+k)(n+k-1)\dots(n+1)}$ cannot become 0, as all $n, k \geq 0$. The smallest energy possible, the ground-state energy, is given by

$$E_{\min} = \hbar\omega(0+\frac{1}{2}) = \frac{\hbar\omega}{2} \; . \tag{1.60}$$

This is the famous "zero-point energy" that is > 0 even in the lowest possible state.

1.5.2 Eigenstates and non-degeneracy

In the previous section, we have found the eigenenergies by only using operators a and $a^\dagger$ and their commutator $[a, a^\dagger] = 1$, which was in turn inferred from the canonical commutation relation $[p, x] = -i\hbar 1$. We did there not have to specify whether the quantum states were functions of position space, momentum space or anything else.

In order to explicitly calculate an eigenvector, we now decide to work in position space. [6] One way would be to solve the differential equation (1.48) for specific energies $E_n = \hbar\omega(n + \frac{1}{2})$. A simpler equation is obtained by noting that applying the lowering operator a to the ground state $|n = 0\rangle$ has to give 0 as a result, see the discussion of the non-negative eigenvalues above. In position space, this gives

$$a|n=0\rangle = \frac{1}{\sqrt{2}}(x + \frac{i}{\hbar}p)|n=0\rangle = 0 \quad \Rightarrow \quad \frac{1}{\sqrt{2}}(x + \partial_x)\psi_{n=0}(x) = 0\ . \tag{1.61}$$

This differential equation is solved by $\psi_{n=0}(x) = C\mathrm{e}^{-x^2/2}$ and the only [7] constant C is found be requiring the eigenstate to be normalized to 1:

$$\psi_{n=0}(x) = \frac{1}{\sqrt[4]{\pi}}\mathrm{e}^{-x^2/2} \tag{1.62}$$

The ground state is thus uniquely determined and non-degenerate.

As we have seen, applying $a^\dagger$ to an eigenstate $|n\rangle$ gives an eigenstate $|n+1\rangle$. However, might there be *another* eigenstate to the eigenvalue $n+1$, denoted by $|\widetilde{n+1}\rangle$? We have shown in Sec. 1.5.1 that applying a to any eigenstate for eigenvalue $n+1$ (now assumed to be degenerate) will give us an eigenstate for eigenvalue n.

$$a|n+1\rangle = \sqrt{n+1}|n\rangle, \tag{1.63}$$
$$a|\widetilde{n+1}\rangle = \sqrt{n+1}|\tilde{n}\rangle, \tag{1.64}$$
$$\dots$$

If the eigenvalue n is non-degenerate (as is true at least for $n = 0$), $|n\rangle$ and $|\tilde{n}\rangle$ can at most differ by a complex phase $|\tilde{n}\rangle = \mathrm{e}^{i\phi}|n\rangle$. Going back up to $n+1$, we have on one hand

$$a^\dagger|\tilde{n}\rangle = a^\dagger\frac{1}{\sqrt{n+1}}a|\widetilde{n+1}\rangle = \frac{n+1}{\sqrt{n+1}}|\widetilde{n+1}\rangle = \sqrt{n+1}|\widetilde{n+1}\rangle \quad \text{and}$$
$$a^\dagger|n\rangle = a^\dagger\frac{1}{\sqrt{n+1}}a|n+1\rangle = \sqrt{n+1}|n+1\rangle\ . \tag{1.65}$$

On the other hand,

$$a^\dagger|\tilde{n}\rangle = a^\dagger\mathrm{e}^{i\phi}|n\rangle = \mathrm{e}^{i\phi}\sqrt{n+1}|n+1\rangle\ . \tag{1.66}$$

$|\widetilde{n+1}\rangle$ and $|n+1\rangle$ thus also at most differ by a complex phase, as $|\widetilde{n+1}\rangle = \mathrm{e}^{i\phi}|n+1\rangle$, and eigenvalue $n+1$ can not have more than one linearly independent eigenvector, meaning that it is not degenerate. Since the ground state is non-degenerate, all eigenstates are non-degenerate.

[6] Due to the equivalent role of x and p played in the Hamiltonian, momentum space would work equally well.

[7] As the differential equation is of first order, we expect only one integration constant in the general solution.

1.5.3 Coherent States

To practice some concepts from both the Harmonic oscillator and time dependence, we look at so-called coherent states. They were originally introduced as quantum states most closely related to what one would expect from classical harmonic oscillators and have since become useful in situations where "phase coherence" is important. As we are going to see, their similarity to classical solutions makes these states are rather 'wave like'. In contrast to the quite 'particle-like' eigenstates of the harmonic oscillator, this makes them a useful choice whenever wave aspects of quantum systems are important, e.g., phase coherence in a laser.

We know of course that positon and momentum of the *classical* harmonic oscillator, oscillate with a cosine in time. Is we want to mimic such bahavior for a "calssical and wave-like" state, we would naturally focus on the expectation values in the quantum mechanical case. However, these vanish for all eigenstates of the harmonic osciallator. We are here going to construct states where $\langle\hat{x}\rangle$ and $\langle\hat{p}\rangle$ do not vanish and instead oscillate like in the classical state.

Expectation values for $\hat{p}$ and $\hat{x}$ can be obtained from those of a and $a^\dagger$ as

$$\langle\psi|\hat{x}|\psi\rangle = \frac{1}{\sqrt{2}}\langle\psi|a+a^\dagger|\psi\rangle = \frac{1}{\sqrt{2}}\big(\langle\psi|a\psi\rangle + \underbrace{\langle a\psi|\psi\rangle}_{=\langle\psi|a\psi\rangle^*}\big) = \sqrt{2}\mathcal{R}(\langle\psi|a|\psi\rangle) \quad \text{and} \tag{1.67}$$

$$\langle\psi|\hat{p}|\psi\rangle = \frac{-i\hbar}{\sqrt{2}}\langle\psi|a-a^\dagger|\psi\rangle = \frac{-i\hbar}{\sqrt{2}}\big(\langle\psi|a\psi\rangle - \langle a\psi|\psi\rangle\big) = \sqrt{2}\hbar\mathcal{I}(\langle\psi|a|\psi\rangle)\,. \tag{1.68}$$

If $|\psi\rangle$ are eigenstates of the harmonic oscillator, we can read off immediately that $\langle\hat{x}\rangle$ and $\langle\hat{p}\rangle$ vanish.

Before looking for alternative states to consider, we investigate the time dependence of $\langle\psi|a|\psi\rangle$:

$$\begin{aligned}\frac{d}{dt}\langle\psi(t)|a|\psi(t)\rangle &= \langle\frac{d}{dt}\psi(t)|a|\psi(t)\rangle + \langle\psi(t)|a\frac{d}{dt}|\psi(t)\rangle = \frac{i}{\hbar}\langle H\psi(t)|a|\psi(t)\rangle - \frac{i}{\hbar}\langle\psi(t)|aH|\psi(t)\rangle = \\ &= i\langle\psi(t)|[H,a]|\psi(t)\rangle = i\omega\langle\psi(t)|[a^\dagger a,a]|\psi(t)\rangle = i\omega\langle\psi(t)|\big(\underbrace{[a^\dagger,a]}_{=-1}a + a^\dagger\underbrace{[a,a]}_{=0}|\psi(t)\big)\rangle = \\ &= -i\omega\langle\psi(t)|a|\psi(t)\rangle\,,\end{aligned} \tag{1.69}$$

which can be integrated to

$$\langle\psi(t)|a|\psi(t)\rangle = \mathrm{e}^{-i\omega t}\langle\psi(t=0)|a|\psi(t=0)\rangle\,. \tag{1.70}$$

Two things are here to note:

1. As classical way to solve the classical harmonic oscillator starts with combining real p and q into a complex varaible $z(t) = \omega q(t) + \frac{i}{m}p(t)$, so that their two differential equations are also combined into one: With $\dot{q} = \frac{p}{m}$, $\dot{p} = -kq$ and $\omega = \sqrt{k/m}$, one finds $\dot{z} = -i\omega z$, in striking similarity to Eq. (1.69) This is of course solved by $z(t) = z_0\mathrm{e}^{-i\omega t}$, compare with (1.70). The original quantities are then $q(t) = \mathcal{R}(z)/\omega$ and $p(t) = m\mathcal{I}(z)$, which we analogously find in Eq. (1.68).
2. If we want to mimic the classical solution, eigenstates of a would be handy: If $|\psi(t)\rangle$ is an eigenstate of a, then applying a does not move any weight into orthogoinal states (as

happens with $|n\rangle$) and the scalar product in Eq. (1.68) becomes the maximal possible value possible for normalized states. The correct time evolution of the expectation values would be ensured by Eq. (1.69).

Even though a is not Hermitian, right eigenstates exist and are given by

$$|z\rangle = \mathrm{e}^{-\frac{|z|^2}{2}} \sum_{n=0}^{\infty} \frac{z^n}{\sqrt{n!}} |n\rangle = \mathrm{e}^{-\frac{|z|^2}{2}} \mathrm{e}^{a^\dagger z} |n=0\rangle \ , \tag{1.71}$$

where z is an arbitrary complex number. To see that they are indeed eigenstates, apply a

$$\begin{aligned} a|z\rangle &= \mathrm{e}^{-\frac{|z|^2}{2}} \sum_{n=0}^{\infty} \frac{z^n}{\sqrt{n!}} \underbrace{a|n\rangle}_{=\sqrt{n}|n-1\rangle} = \mathrm{e}^{-\frac{|z|^2}{2}} \sum_{m=-1}^{\infty} \frac{z^{m+1}}{\sqrt{(m+1)!}} \underbrace{\sqrt{m+1}}_{=0 \text{ for } m=-1} |m\rangle = \\ &= \mathrm{e}^{-\frac{|z|^2}{2}} \sum_{m=0}^{\infty} \frac{z^{m+1}}{\sqrt{(m+1)!}} \sqrt{m+1}|m\rangle = \mathrm{e}^{-\frac{|z|^2}{2}} z \sum_{m=0}^{\infty} \frac{z^m}{\sqrt{m!}} |m\rangle = z|z\rangle \ . \end{aligned} \tag{1.72}$$

Starting at $t = 0$ from $z(t = 0) = z$, the expectation value for $t \neq 0$ is found from Eq. (1.70) as

$$\langle z(t)|a|z(t)\rangle = \mathrm{e}^{-i\omega t} \langle z(t=0)|a|z(t=0)\rangle = \mathrm{e}^{-i\omega t} z \tag{1.73}$$

and indeed rotates like the classical $z(t) = \mathrm{e}^{-i\omega t} z_0$. Analogously, expectation values $\langle z|\hat{x}|z\rangle$ and $\langle z|\hat{p}|z\rangle$ behave like classical $x(t)$ and $p(t)$, e.g. the choice of real z at $t = 0$ leads to

$$\begin{aligned} \langle \hat{x} \rangle_z &= \sqrt{2}\, \mathcal{R}(\mathrm{e}^{i\omega t} z) = \sqrt{2} z \cos \omega t \quad \text{and} \\ \langle \hat{p} \rangle_z &= \sqrt{2}\hbar\, \mathcal{I}(\mathrm{e}^{i\omega t} z) = \sqrt{2}\hbar z \sin \omega t. \end{aligned} \tag{1.74}$$

For arbitrary $z = x_0 + \frac{i}{\hbar} p_0 = z_0 \mathrm{e}^{i\phi}$, one finds

$$\begin{aligned} \langle \hat{x} \rangle_z &= \sqrt{2}(x_0 \cos \omega t - \frac{p_0}{\hbar} \sin \omega t) = \sqrt{2} z_0 (\cos \phi_0 \cos \omega t - \sin \phi_0 \sin \omega t) = \sqrt{2} z_0 \cos(\omega t + \phi_0) \\ \langle \hat{p} \rangle_z &= \sqrt{2}\hbar (x_0 \sin \omega t + \frac{p_0}{\hbar} \cos \omega t) = \sqrt{2}\hbar z_0 (\cos \phi_0 \sin \omega t + \sin \phi_0 \cos \omega t) \\ &= \sqrt{2}\hbar z_0 \sin(\omega t + \phi_0) \ . \end{aligned} \tag{1.75}$$

This phase ϕ_0 that has been introduced here is very different from the arbitrary und unmeasurable phase that can be assigned to any quantum state. As it is given by physical observables x_0 and p_0, which can in principle be measured, it correponds to a physical quantity. This phase can be used to define concepts like phase coherence.

At $t = 0$, the expectation value of the energy is given by

$$\langle z|H|z\rangle = \hbar\omega \langle z|a^\dagger a + \frac{1}{2}|z\rangle = \hbar\omega \left(||a|z\rangle||^2 + \frac{1}{2} \right) = \hbar\omega \left(|z|^2 \cdot + \frac{1}{2} \right) \tag{1.76}$$

and does not depend on time, because H commutes with the time-evolution operator. For large z, the constant $\frac{1}{2}$ is negligible and $E \approx \hbar\omega |z|^2$, so that $|z| = z_0$ can be seen as the analogue of the wave's amplitude. We have here consequently a fairly classical behaviour characterized by an amplitude z_0 and a phase ϕ_0.

Let us finally discuss the uncertainty of momentum and position when measured in a coherent state.

$$\begin{aligned}\langle \hat{x}^2\rangle_t &= \frac{1}{2}\langle z|(a_H(t)+a_H^\dagger(t))^2|z\rangle = \frac{1}{2}\Bigg(\underbrace{\langle z|a_H(t)^2|z\rangle}_{=z^2\mathrm{e}^{-2i\omega t}} + \underbrace{\langle z|a_H^\dagger(t)^2|z\rangle}_{=(z^*)^2\mathrm{e}^{2i\omega t}} + \\ &\quad + \underbrace{\langle z|a_H(t)a_H^\dagger(t)|z\rangle}_{=\langle z|a_H^\dagger a_H|z\rangle+1=|z|^2+1} + \underbrace{\langle z|a_H^\dagger(t)a_H(t)|z\rangle}_{=|z|^2}\Bigg) \\ &= \frac{1}{2} + \frac{1}{2}\left(z\mathrm{e}^{-i\omega t}+z^*\mathrm{e}^{i\omega t}\right)^2 \quad \text{and} \qquad (1.77)\end{aligned}$$

$$\langle \hat{p}^2\rangle_t = \frac{-\hbar^2}{2}\langle z|(a_H(t)-a_H^\dagger(t))^2|z\rangle = \cdots = \frac{\hbar^2}{2} - \frac{\hbar^2}{2}\left(z\mathrm{e}^{-i\omega t}-z^*\mathrm{e}^{i\omega t}\right)^2 \qquad (1.78)$$

Subtracting the squares of $\langle\hat{x}(t)\rangle$ and $\langle\hat{p}(t)\rangle$ leaves only the constant and yields

$$\sqrt{\langle\hat{x}^2\rangle-\langle\hat{x}\rangle^2}\cdot\sqrt{\hat{p}^2\rangle-\langle\hat{p}\rangle^2} = \frac{1}{\sqrt{2}}\cdot\frac{\hbar}{\sqrt{2}} = \frac{\hbar}{2}\,. \qquad (1.79)$$

Even coherent states with a large amplitude – and consequently a large energy – thus minimize the Heisenberg uncertainty, in contrast to eigenstates $|n\rangle$, where it grows with n. Again, this illustrates that coherent states are as similar to a classical trajectory as quantum states can possibly be.

We have here seen that expectation values of some operators (like x and p) can *happen* to be constant for some initial states (e.g. $|n\rangle$), but evolve for other choices (e.g. $|z\rangle$). In contrast, the expectation value of the Hamiltonian remains constant for arbitrary initial states. This second situation will be explored in more detail in the next section on symmetries and conserved quantities.

2 Symmetries and conserved quantities

Le Bellac, Sakurai I

When a Hamiltonian is invariant with respect to some symmetry, then this symmetry is also preserved under time evolution. A state that has it will thus keep it. Examples can be discrete symmetries like parity (i.e., even/odd wave functions in 1D) or continuous symmetries, which are connected to an observable that is conserved. As the symmetries discussed here are expressed in space, we will mostly use states written in position space.

In describing symmetries, we use *unitary operators*: A unitary operator is one that keeps the scalar product (i.e. norms of states and "angles" between them) invariant, and we thus ask that $U^{-1} = U^\dagger$, see Sec. 1.1.2. This means that the relations between elements of the Hilbert space are not changed by the operator, which seem desirable for a symmetry. [1] For an operator to be invariant with respect to a symmetry encoded in U means that the operator must not change if all vectors are transformed using U, e.g., a rotationally invariant Hamiltonian should not change if we rotate the universe. Accordingly,

$$\langle\phi|U^\dagger H U|\psi\rangle = \langle\phi|U^{-1} H U|\psi\rangle = \langle\phi|H|\psi\rangle \tag{2.1}$$

for *all* ϕ and ψ and thus

$$U^{-1}HU = H, \quad \text{and} \quad HU = UH, \quad \text{resp.} \quad [H,U] = 0\,. \tag{2.2}$$

In the case of a continuous symmetry, unitary U is not Hermitian and thus not itself the conserved observable.

Further, we ask operators describing symmetries to have the following properties:

1. Combining two symmetry operations should give another valid symmetry operation. (If the system is really symmetric w.r.t. to the first operation, it would be weird if the second operator became forbidden.) $T(a)T(b) = T(a \cdot b)$, when repeated, this process should be associative $T(a)T(b)T(c) = T(a)T(b \cdot c) = T(a \cdot b)T(c)$.
2. It should be possible not to change the state at all, i.e., the $\mathbb{I}$-operator should also be a valid symmetry transformation.
3. We should be able to undo a symmetry transformation by another transformation, i.e., there should be an inverse transformation $T^{-1}(a)$ such that $T^{-1}(a)T(a) = T(a)T^{-1}(a) = \mathbb{I}$.

These properties mean that the symmetries form a group. Note that the use of $\cdot$ in the first point is not meant to imply commutativity, i.e., $T(a)T(b)$ does not have to be the same as $T(b)T(a)$.

[1] In fact, *anti*-unitary operators can also have useful applications as symmetries, but this is a bit more specialized and advanced.

For continuous symmetries, it should also make sense to talk about arbitrarily small transformations very close to $\mathbb{I}$. The following form of such an infinitesimal transformation does not contradict the rules above:

$$T(\epsilon) = \mathbb{I} - i\frac{\epsilon}{\hbar}K \ . \tag{2.3}$$

If operator K is chosen to be Hermitian, then $T(\epsilon)$ is unitary, because $T^\dagger(\epsilon) = \mathbb{I} + i\frac{\epsilon}{\hbar}K^\dagger$ and $T^{-1}(\epsilon) = \mathbb{I} + i\frac{\epsilon}{\hbar}K + \mathcal{O}(\epsilon^2)$. For infinitesimal ϵ, orders higher than linear in ϵ can be neglected and $K = K^\dagger$ thus makes $T^\dagger(\epsilon) = T^{-1}(\epsilon)$ and $T(\epsilon)$ becomes unitary. As T is a function of K and commutes with H and as K is (in contrast to T) hermitian, one can here start to suspect that K might be the conserved quantity.

Concerning the group axioms:

1. Combining two transformations gives

$$T(\epsilon)T(\epsilon') = \left(\mathbb{I} - i\frac{\epsilon}{\hbar}K\right)\left(\mathbb{I} - i\frac{\epsilon'}{\hbar}K'\right) = \mathbb{I} - \frac{i}{\hbar}(\epsilon K + \epsilon' K') + \mathcal{O}(\epsilon\epsilon') \tag{2.4}$$

   which becomes for infinitesimal ϵ and ϵ' [i.e., neglecting $\mathcal{O}(\epsilon\epsilon')$] an allowed infinitesimal transformation as $K_{\text{comb.}} = K + \frac{\epsilon'}{\epsilon}K' = K^\dagger_{\text{comb.}}$ is also an "allowed" operator. Finally, the combination of such infinitesimal transformations is also associative.
2. For $\epsilon \to 0$, we obtain the neutral element $\mathbb{I}$.
3. $T(-\epsilon)$ gives the inverse transformation, see the discussion of unitarity above.

A group with such a "generator" K of infinitesimal elements is a Lie group, we are going to explore the consequences of the axioms in more detail for the case of translational invariance.

2.1 Translational invariance

2.1.1 Continuous Translations

Let us define translations via their action on eigenstates $|\vec{r}\rangle$ of the position operator $\hat{\vec{r}}$:

$$T(\vec{a})|\vec{r}\rangle = |\vec{r} + \vec{a}\rangle. \tag{2.5}$$

They fulfill the group axioms, the combination of two translations $T(\vec{b})T(\vec{a})$ into a combined one $T(\vec{a} + \vec{b})$ is moreover commutative, because the addition of real-space vectors is. The inverse element $T^{-1}(\vec{a}) = T(-\vec{a})$. All consideration can likewise be carried out in one or two dimensions. At least when acting on states $|\vec{r}\rangle$, $T(\vec{a})$ clearly preserves the scalar product

$$\langle\vec{r}'|\vec{r}\rangle = \delta(\vec{r} - \vec{r}') \quad \text{and} \quad \langle T(\vec{a})\vec{r}'|T(\vec{a})\vec{r}\rangle = \delta\left((\vec{r} - \vec{a}) - (\vec{r}' - \vec{a})\right) = \delta(\vec{r} - \vec{r}') \ . \tag{2.6}$$

Even without knowing how $T(\vec{a})$ acts on general states $|\phi\rangle$, we can now verify that $T(\vec{a})$ in fact preserves *all* scalar products by inserting twice a $\mathbb{I}$-operator of the form $\mathbb{I} = \int |\vec{r}\rangle\langle\vec{r}|$:

$$\langle T(\vec{a})\phi|T(\vec{a})\psi\rangle = \langle\phi|T^\dagger(\vec{a})T(\vec{a})|\psi\rangle = \int_{\vec{r}} d^3r \int_{\vec{r}'} d^3r' \langle\phi|\vec{r}\rangle \underbrace{\langle\vec{r}|T^\dagger(\vec{a})T(\vec{a})|\vec{r}'\rangle}_{=\delta(\vec{r}-\vec{r}')}\langle\vec{r}'|\psi\rangle =$$

$$= \int_{\vec{r}} d^3r \langle\phi|\vec{r}\rangle\langle\vec{r}|\psi\rangle = \langle\phi|\psi\rangle \ . \tag{2.7}$$

Translations are thus unitary.

We are of course particularly interested in the action of $T(\vec{a})$ on *quantum states*,i.e., on wave functions $\psi(\vec{r}) = \langle\vec{r}|\psi\rangle$. For $|\psi'\rangle = T(\vec{a})|\psi\rangle$, we find

$$\begin{aligned}\psi'(\vec{r}) = \langle\vec{r}|\psi'\rangle = \langle\vec{r}|T(\vec{a})|\psi\rangle = \langle T^\dagger(\vec{a})\vec{r}|\psi\rangle = \langle T^{-1}(\vec{a})\vec{r}|\psi\rangle = \langle T(-\vec{a})\vec{r}|\psi\rangle = \\ = \langle\vec{r}-\vec{a}|\psi\rangle = \psi(\vec{r}-\vec{a}) \ .\end{aligned} \tag{2.8}$$

If this seems at first counterintuitive, one can understand it by considering that when $T(\vec{a})$ moves the position eigenstates $|\vec{r}\rangle$ to the right, a "fixed" wave function ψ moves to the left relative to $\vec{r}$.

In order to find out what the generator K is in the case of translations, we can go to position-space representation and find that

$$T_x(\epsilon)\psi(\vec{r}) = \psi(\vec{r}-\epsilon\mathrm{e}_x) = \psi(\vec{r}) - \epsilon\frac{\partial\psi}{\partial x} = (\mathbb{I} - \frac{i\epsilon}{\hbar}\hat{p}_x)\psi(\vec{r}) \ . \tag{2.9}$$

We can thus set $K = \hat{p}_x$ and find that infinitesimal translations along x are caused by the operator

$$T_x(\epsilon) = \mathbb{I} - \frac{i\epsilon}{\hbar}\hat{p}_x \tag{2.10}$$

The conserved observable K associated with translational invariance is consequently momentum.

For finite translations, we can proceed as in the case of the time-evolution operator

$$\frac{d}{da}T_x(a) = \lim_{\epsilon\to 0}\frac{T_x(a+\epsilon)-T_x(a)}{\epsilon} = \lim_{\epsilon\to 0}\frac{T_x(\epsilon)T_x(a)-T_x(a)}{\epsilon} = \lim_{\epsilon\to 0}\frac{T_x(\epsilon)-\mathbb{I}}{\epsilon}T_x(a) = -\frac{i}{\hbar}\hat{p}_xT_x(a) \tag{2.11}$$

which can be integrated to

$$T_x(a) = \mathrm{e}^{-\frac{ia}{\hbar}\hat{p}_x} \ . \tag{2.12}$$

The transition back to three-dimensional space can be achieved by making both $\epsilon \to \epsilon\vec{a}$ and K vector-valued in Eq. (2.3), this amounts to combining three infinitesimal translations (along x, y and z) into one new infinitesimal translation - which is allowed according to the group axioms. As components of $\vec{r}$ commute with momenta for *other* components and with each other, (infinitesimal and finite) translations along different directions commute and it does not even matter in which order they are combined. For example, Eqs. (2.9) and (2.10) become

$$T(\epsilon\vec{a})\psi(\vec{r}) = \psi(\vec{r}-\epsilon\vec{a}) = \psi(\vec{r}) - \epsilon\vec{a}\nabla\psi = (\mathbb{I} - \frac{i\epsilon}{\hbar}\hat{\vec{p}})\psi(\vec{r}) \tag{2.13}$$

and

$$T(\epsilon\vec{a}) = \mathbb{I} - \frac{i\epsilon}{\hbar}\vec{a}\hat{\vec{p}} \ . \tag{2.14}$$

and finite translations (2.12) are written as

$$T(\vec{a}) = \mathrm{e}^{-\frac{i}{\hbar}\vec{a}\hat{\vec{p}}} \,. \tag{2.15}$$

One notices here that this finite translation by and arbitrary vector $\vec{a}$ can be written as the product of finite translations along the three components of the vector, i.e.,

$$T(\vec{a}) = \mathrm{e}^{-\frac{i}{\hbar}a_x\hat{p}_x} \cdot \mathrm{e}^{-\frac{i}{\hbar}a_y\hat{p}_y} \cdot \mathrm{e}^{-\frac{i}{\hbar}a_z\hat{p}_z} \,. \tag{2.16}$$

Translations combine in this straightforward manner $T(\vec{a}_1+\vec{a}_2) = T(\vec{a}_1)T(\vec{a}_2)$ because they commute, resp., their generators, components of $\hat{\vec{p}}$, commute $[\hat{p}_\alpha, \hat{p}_\beta] = 0$.

2.1.2 Translations by lattice vectors

Suppose a particle moving in a one-dimensional [2] periodic potential with a lattice constant a, i.e., $V(x+a) = V(x)$ for all x. Let us further assume periodic boundary conditions, i.e., we ask that the wave function should be periodic with a period Na. This is a choice, it would similarly be valid to ask for, say, antiperiodic boundary conditions with $\psi(x+Na) = -\psi(x)$, or we could alternativel deal with the full infinite space as one does for the free particle. [3] The choice of periodic boundary conditions is often made in solid-state physics.

The Hamiltonian describing such a system is not translationally invariant (as there is a potential varying in x), but should be symmetric with respect to moving the particle by a, i.e., $T(-a)HT(a) = H$. As infinitesimal translations are here not an allowed symmetry operation, we can here not define a generator. Unitarity and group structure still remain.

As an example, we discuss a very simple Hamiltonian having these properties, the one-dimensional "tight-binding" Hamiltonian. It describes a particle that can sit in any of N boxes of a periodic potential, but is not concerned with the form of the wave function within a box. The state $|m\rangle$ of the particle is thus fully described by the index m of the box it occupies and the periodic boundary conditions require $|m+N\rangle = |m\rangle$. The potential barrier between the boxes is finite and the particle can tunnel from one box to the next. [4] The Hamiltonian describing this system can be written as

$$\begin{aligned} \langle m|H|m\rangle &= H_{m,m} = \epsilon_0 \\ \langle m-1|H|m\rangle &= H_{m-1,m} = -t \\ \langle m+1|H|m\rangle &= H_{m+1,m} = -t \\ \langle n|H|m\rangle &= 0 \qquad\qquad \text{for } |n-m| > 1. \end{aligned} \tag{2.17}$$

Where the periodic boundary conditions must be taken into account properly.

The states $|m\rangle$ are not eigenstates of H, as $H|m\rangle$ has components along $|m\pm 1\rangle$ as well. However, as H commutes with a translation by a, we can make use of this symmetry in finding the eigenstates. In looking for eigenstates of $T(a)$, analogy with the momentum eigenstate in position representation motivates to following Ansatz

$$|k\rangle = \frac{1}{\sqrt{N}} \sum_{m=1}^{N} \mathrm{e}^{ikm}|m\rangle \,. \tag{2.18}$$

[2] Considerations in higher dimensions are analogous.

[3] For very large systems $N \to \infty$, the differences between the three choices become fortunately very small.

[4] It is also assumed that tunneling is so small that only connect nearest-neighbor boxes, a condition that can be relaxed.

We measure distances m ununits of a and momenta k accordingly in $\frac{1}{a}$. We check that it is indeed an eigenstate of $T(a)$ as above

$$T(a)|k\rangle = \frac{1}{\sqrt{N}}\sum_m \mathrm{e}^{ikm}T(a)|m\rangle = \frac{1}{\sqrt{N}}\sum_{m=1}^{N}\mathrm{e}^{ikm}|m+1\rangle = \frac{1}{\sqrt{N}}\sum_{m'=2}^{N+1}\mathrm{e}^{ikm'}\mathrm{e}^{-ik}|m'\rangle =$$
$$= \frac{1}{\sqrt{N}}\sum_{m'=1}^{N}\mathrm{e}^{ikm'}\mathrm{e}^{-ik}|m'\rangle = \mathrm{e}^{-ik}|k\rangle\ . \tag{2.19}$$

We have here used periodic boundary conditions when going from the sum over m (running from 1 to N) to a sum over $m' = m+1$ (running from 2 to $N+1 = 1$). They ensure that all states are still part of the sum. Periodic boundary conditions require (e.g. for this trick to work) that $\mathrm{e}^{ik(N+a)} = \mathrm{e}^{ika}$, i.e., that $kN = 2\pi l$ (with integer l). They consequently lead to a quantization of momentum to integer multiples of $2\pi/N$.

Given that $T(a)$ leaves H invariant, they share an eigensystem and $|k\rangle$ should be eigenstates of H. This can be checked, by inserting a $\mathbb{I} = \sum_n |n\rangle\langle n|$ and using the same relabeling of the summation index as above:

$$H|k\rangle = \frac{1}{\sqrt{N}}\sum_m \mathrm{e}^{ikm}H|m\rangle = \frac{1}{\sqrt{N}}\sum_m\sum_n \mathrm{e}^{ikm}|n\rangle\underbrace{\langle n|H|m\rangle}_{\text{see (2.17)}} =$$
$$= \frac{1}{\sqrt{N}}\sum_m \mathrm{e}^{ikm}|m\rangle\epsilon_0 + \frac{1}{\sqrt{N}}\sum_m \mathrm{e}^{ikm}|m-1\rangle(-t) + \frac{1}{\sqrt{N}}\sum_m \mathrm{e}^{ikm}|m+1\rangle(-t) =$$
$$= \epsilon_0|k\rangle - \frac{t}{\sqrt{N}}\underbrace{(\mathrm{e}^{ik}+\mathrm{e}^{-ik})}_{2\cos k}\sum_m \mathrm{e}^{ikm}|m\rangle = (\epsilon_0 - 2t\cos k)|k\rangle \tag{2.20}$$

$|k\rangle$ is indeed an eigenstate of H, with eigenvalue $\epsilon(k) = \epsilon_0 - 2t\cos k$.

From the theory of Fourier transforms, we know that $|k\rangle$ with $k \in [0, 2\pi[$ are a basis and thus give a complete set of eigenvectors. The eigenvalues are almost all (except for $k = 0$ and $k = \pi$) doubly degenerate. Usually the "band structure" $\epsilon(k)$ is plotted for the shifted (but equivalent) interval $]-\pi, \pi]$ instead of $[0, 2\pi[$. The conserved quantity k is not the usual momentum (given by $-i\hbar\partial x$), because infinitesimal translations do not leave the Hamiltonian invariant. Instead, k is the so-called "crystal momentum".

A Hamiltonian like Eq. (2.17) is justified when the tunneling $|t|$ between potential wells is much smaller than the energy spacing between internal states of the well. An example would be tightly bound electrons in fairly localized orbitals of a solid. When going from isolated atoms to atoms coupled by t, this level at energy ϵ_0 is broadened into an energy band with width $2t$.

2.2 Rotations

Another important symmetry is symmetry with respect to rotations. Examples for rotationally invariant potentials are any that only depend on distance $|\vec{r}|$, but not on the direction of $\vec{r}$. Rotations can be combined according to group axioms (there are also inverse and neutral elements), but rotations around different axes do not generally commute, i.e., the group is not commutative.

As the notion of infinitesimal rotations makes sense, we have a Lie group and write according to (2.3) for an infinitesimal rotation around the z-axis

$$R_z(d\alpha) = \mathbb{I} - \frac{id\alpha}{\hbar}\hat{G} \,. \tag{2.21}$$

In order to find out what $\hat{G}$ is, it is here easiest to go to position space. Analogously to translations, we can deduce that wave functions rotate in the way "opposite" to the eigenstates of $\hat{\vec{r}}$:

$$\begin{aligned} R_z(\alpha)\psi(\vec{r}) = \langle\vec{r}|R_z(\alpha)|\psi\rangle = \langle R_z^\dagger(\alpha)\vec{r}|\psi\rangle = \langle R_z^{-1}(\alpha)\vec{r}|\psi\rangle = \langle R_z(-\alpha)\vec{r}|\psi\rangle = \\ = \psi(R_z(-\alpha)\vec{r}) \end{aligned} \tag{2.22}$$

and for infinitesimal rotations

$$\begin{aligned} R_z(d\alpha)\psi(\vec{r}) = \psi(R_z(-d\alpha)\vec{r}) = \psi(x-(-d\alpha)y, y+(-d\alpha)x, z) = \\ = \psi(\vec{r}) - d\alpha(x\partial_y\psi - y\partial_x\psi) + \mathcal{O}(d\alpha^2) \end{aligned} \tag{2.23}$$

Comparison to the orbital angular momentum $\hat{\vec{L}} = \hat{\vec{r}} \times \hat{\vec{p}} = -i\hbar\vec{r} \times \nabla$ leads to

$$R_z(d\alpha)\psi(\vec{r}) = (\mathbb{I} - \frac{id\alpha}{\hbar}L_z)\psi(\vec{r}) \,. \tag{2.24}$$

2.2.1 Finite Rotations around arbitrary axes

In order to address finite rotations around arbitrary axes, we make again use of the fact that we can move the transformation of the quantum state onto the real-space vector $\vec{r}$.

$$\begin{aligned} R_{\vec{n}}(\alpha)\psi(\vec{r}) = \langle\vec{r}|R_{\vec{n}}(\alpha)|\psi\rangle = \langle R_{\vec{n}}^\dagger(\alpha)\vec{r}|\psi\rangle = \langle R_{\vec{n}}^{-1}(\alpha)\vec{r}|\psi\rangle = \langle R_{\vec{n}}(-\alpha)\vec{r}|\psi\rangle = \\ = \psi(R_{\vec{n}}(-\alpha)\vec{r}) \,. \end{aligned} \tag{2.25}$$

The advantage is that rotations of vectors $\vec{r}$ are accessible to geometry. When speaking about rotations in three-dimensional space, there are two useful ways of writing them. The first describes a rotation by an axis $\vec{n}$, a vector of unit length, and the angle of the rotation around this axis α. The second way describes a rotation as a 3×3 orthogonal matrix; its rows are given by orthogonal unit vectors defining a coordinate system: μ, $\nu \perp \mu$ and $\nu \times \mu$. The advantage of this second approach is that combined rotations are given by the matrix product of the respective individual rotation matrices. In both descriptions, three parameters are needed.

We are here going to discuss rotations in terms of 3×3 matrices and are going to find them by going via infinitesimal rotations around a given axis. As rotations fulfill all the requirements for a Lie group, we expect to be able to write them as

$$R_{\vec{n}}(\alpha)\vec{r} = \mathrm{e}^{-i\alpha T_{\vec{n}}}\vec{r} \qquad \text{and} \quad R_{\vec{n}}(d\alpha)\vec{r} = (\mathbb{I} - id\alpha T_{\vec{n}})\vec{r} \,, \tag{2.26}$$

where $T_{\vec{n}}$ is some hermitian operator (i.e., a 3×3 matrix) acting on $\vec{r}$ that we still have to identify.

In order to do so, we investigate the geometry of an infinitesimal rotation $d\alpha$ of vector $\vec{r}$ around an axis $\vec{n}$, with $|\vec{n}| = 1$:

- The component $\vec{r}_\parallel = (\vec{r}\cdot\vec{n})\vec{n}$ remains unchanged.
- The remaining component $\vec{r}_\perp = \vec{r} - (\vec{r}\cdot\vec{n})\vec{n}$ lies in the plane orthogonal to $\vec{n}$. As basis vectors of this plane, we choose
  - $p_1 = \vec{r}_\perp/|\vec{r}_\perp|$ and
  - the vector orthogonal to both $\vec{n}$ and p_1, i.e., $p_2 = \vec{n}\times\vec{r}_\perp/|\vec{r}_\perp| = \vec{n}\times\vec{r}/|\vec{r}_\perp|$.
- The rotation of $\vec{r}_\perp$:
  - Before the rotation, $\vec{r}$ has a component $|\vec{r}_\perp|$ along p_1 and 0 along p_2.
  - After a rotation by α, the new vector will have a component $|\vec{r}_\perp|\cos\alpha$ along p_1 and $|\vec{r}_\perp|\sin\alpha$ along p_2.

These steps are summarized in the equation

$$\begin{aligned}\vec{r}' = R_{\vec{n}}(\alpha)\vec{r} &= \underbrace{(\vec{r}\cdot\vec{n})\vec{n}}_{\text{unchanged}} + |\vec{r}_\perp|\cos\alpha\frac{\vec{r}_\perp}{|\vec{r}_\perp|} + |\vec{r}_\perp|\sin\alpha\frac{\vec{n}\times\vec{r}}{|\vec{r}_\perp|} = \\ &= (\vec{r}\cdot\vec{n})\vec{n} + \big(\vec{r} - (\vec{r}\cdot\vec{n})\vec{n}\big)\cos\alpha + \vec{n}\times\vec{r}\,\sin\alpha\,. \end{aligned} \tag{2.27}$$

For infinitesimal $d\alpha$ and neglecting higher orders, one can use $\sin d\alpha \approx d\alpha$ and $\cos d\alpha \approx 1$ and thus

$$\vec{r}' = R_{\vec{n}}(d\alpha)\vec{r} = \vec{r} + d\alpha\underbrace{\vec{n}\times}_{-iT_{\vec{n}}}\vec{r}\,, \tag{2.28}$$

where the identification with $-iT_{\vec{n}}$ comes from comparison with (2.26). Written using the individual components x, y and z of $\vec{r}$, one gets the matrix representation of $-iT_{\vec{n}}$ as

$$-iT_{\vec{n}}\,\vec{r} = \vec{n}\times\vec{r} = \begin{pmatrix} n_y z - n_z y \\ n_z x - n_x z \\ n_x y - n_y x \end{pmatrix} = \begin{pmatrix} 0 & -n_z & n_y \\ n_z & 0 & -n_x \\ -n_y & n_x & 0 \end{pmatrix}\vec{r}\,. \tag{2.29}$$

Introducing three matrices

$$T_x = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \quad T_y = \begin{pmatrix} 0 & 0 & i \\ 0 & 0 & 0 \\ -i & 0 & 0 \end{pmatrix}, \quad \text{and} \quad T_z = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{2.30}$$

and collecting them into a vector $\vec{T} = (T_x, T_y, T_z)$ of matrices, we can write $T_{\vec{n}}$ as $\vec{n}\vec{T}$ and

$$R_{\vec{n}}(d\alpha) = \mathbb{I} - i\,d\alpha\,\vec{n}\vec{T}\,, \tag{2.31}$$

which can be integrated to

$$R_{\vec{n}}(\alpha) = \mathrm{e}^{i\,\alpha\,\vec{n}\vec{T}}\,. \tag{2.32}$$

The integration is analogous to the previous cases and makes use of the fact that $R_{\vec{n}}(\alpha + d\alpha) = R_{\vec{n}}(d\alpha)R_{\vec{n}}(\alpha)$ and that all 3D-rotations around one axis commute.

To carry over the rotation to wave functions, we proceed as with the case of L_z, only writing the expansion in all three coordinates

$$R_{\vec{n}}(d\alpha)\psi(\vec{r}) = \psi(R_{\vec{n}}(-d\alpha)\vec{r}) = \psi(\vec{r} + i\ d\alpha\ \vec{n}\vec{T}\vec{r}) \approx \psi(\vec{r}) + i\ d\alpha \left[\frac{\partial\psi}{\partial x}(\vec{n}\vec{T}\vec{r})_x + \frac{\partial\psi}{\partial y}(\vec{n}\vec{T}\vec{r})_y + \frac{\partial\psi}{\partial z}(\vec{n}\vec{T}\vec{r})_z\right] \tag{2.33}$$

$(\vec{n}\vec{T}\vec{r})_\gamma$ is here the component $\gamma = x, y, z$ of $(\vec{n}\vec{T}\vec{r})$. Replacing the components by using Eq.(2.29), the rotated wave function becomes

$$\begin{aligned} R_{\vec{n}}(d\alpha)\psi(\vec{r}) &\approx \psi(\vec{r}) + i\ d\alpha \left[i\frac{\partial\psi}{\partial x}(n_y z - n_z y) + i\frac{\partial\psi}{\partial y}(n_z x - n_x z) + i\frac{\partial\psi}{\partial z}(n_x y - n_y x)\right] = \\ &= \psi(\vec{r}) + i\ d\alpha\Big[n_x(\underbrace{i\frac{\partial\psi}{\partial z}}_{=-\frac{\hat{p}_z}{\hbar}\psi}\ y - i\frac{\partial\psi}{\partial y}z) + n_y(i\frac{\partial\psi}{\partial x}z - i\frac{\partial\psi}{\partial z}x) + n_z(i\frac{\partial\psi}{\partial y}x - i\frac{\partial\psi}{\partial x}y)\Big] = \\ &= \psi(\vec{r}) - i\frac{d\alpha}{\hbar}\left[n_x L_x + n_y L_y + n_z L_z\right]\psi(\vec{r}) = \left(\mathbb{I} - i\frac{d\alpha}{\hbar}\vec{n}\hat{\vec{L}}\right)\psi(\vec{r})\ . \end{aligned} \tag{2.34}$$

Eq. (2.24) thus generalizes in the "expected" way, $L_z \to \vec{n}\vec{L}$, where role of the angular momentum along an axis $\vec{n}$ is taken by $\vec{n}\vec{L}$, as in classical physics. [5] As several rotations around one axis always commute [6], the integration to finite rotations is straightforward

$$R_{\vec{n}}(\alpha) = \mathrm{e}^{-i\frac{\alpha}{\hbar}\vec{n}\hat{\vec{L}}}\ . \tag{2.35}$$

While this result for general rotations makes intuitive sense and illustrates that a general rotation can be built up from "elementary" rotations around the x-, y- and z axes, it is worth noting that combined rotations are not just obtained by adding the axes (or the angles), the situation is thus not as straightforward as with translations, see (2.16).

2.2.2 Generators and Commutation Relations

Equations (2.32) describing rotations of 3D vectors $\vec{r}$ and Eq. (2.35) describing rotations of wave functions clearly have analogous structure. When going from rotations around, say, the z-axis to rotations around general axes, the important part is how, exactly, several rotations (e.g. around x-, y-, and z-axes) combine into a single one. This informations is encoded in the commutators of the generators $T_{x,y,z}$ resp. $L_{x,y,z}$.

The particularly simple way of adding translations (2.16) arises because generators commute $[\hat{p}_\alpha, \hat{p}_\beta] = 0$. The essentials of the group of rotations SO(3) are encoded in the commutation relations of the matrices T_α, which can be obtained from Eq.(2.30):

$$[T_x, T_y] = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & 0 & i \\ 0 & 0 & 0 \\ -i & 0 & 0 \end{pmatrix} - \begin{pmatrix} 0 & 0 & i \\ 0 & 0 & 0 \\ -i & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix} =$$

[5] It was thus not just an accident that the operator in Eq. (2.24) looks like $\hat{L}_z$.

[6] Or, in operator language, as operator $\vec{n}\hat{\vec{L}}$ necessarily commutes with itself.

$$= \begin{pmatrix} 0 & 0 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} - \begin{pmatrix} 0 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = iT_z \tag{2.36}$$

and analogously for the other combinations. They can be summarized to

$$[T_\alpha, T_\beta] = i\epsilon_{\alpha\beta\gamma} T_\gamma \ , \tag{2.37}$$

where the Levi-Civita tensor ϵ is a fully antisymmetric tensor of rank three, i.e., $\epsilon_{\alpha\beta\gamma} = -\epsilon_{\beta\alpha\gamma} = -\epsilon_{\alpha\gamma\beta}$ and consequently vanishes if any two indices are the same. $\epsilon_{xyz} = 1$ and summation over repeated indices is often implied when it is used (it is here).

Rotations of wave functions are governed by the very similar commutation relations of (orbital) angular momentum:

$$\begin{aligned} [L_x, L_y] &= [yp_z - zp_y, zp_x - xp_z] = [yp_z, zp_x] - \underbrace{[yp_z, xp_z]}_{=0} - \underbrace{[zp_y, zp_x]}_{=0} + [zp_y, xp_z] = \\ &= y[p_z, zp_x] + \underbrace{[y, zp_x]}_{=0} p_z + z\underbrace{[p_y, xp_z]}_{=0} + [z, xp_z]p_y = \\ &= y\underbrace{[p_z, z]}_{=-i\hbar} p_x + yz\underbrace{[p_z, p_x]}_{=0} + x\underbrace{[z, p_z]}_{=i\hbar} p_y + \underbrace{[z, x]}_{=0} p_z p_y = i\hbar(xp_y - yp_x) = i\hbar L_z \ , \end{aligned} \tag{2.38}$$

becoming

$$[L_\alpha, L_\beta] = i\hbar\epsilon_{\alpha\beta\gamma} L_\gamma \ . \tag{2.39}$$

The only difference is the prefactor $\hbar$, which can be removed by going to units of $\hbar$ for $\vec{L}$, see the next chapter. The Lie algebras of rotations of vectors and wave functions are thus equivalent, which implies that the group of rotations is equivalent for both, at least "close to" the identity $\mathbb{I}$. This should not surprise us, because we obtained the rotations operators for wave functions from those of vectors. In the next chapter, however, we will discuss a third instance of the same commutations relations, which describes rotations in a slightly different way.

2.3 Hand waving summary

We have here investigated two examples (translations and rotations) for continuous and one (lattice translations) for discrete symmetries; other discrete symmetries that we (are going to) encounter(ed) are inversion symmetry (parity) and particle exchange. Both for continuous and discrete symmetries, 'something' is conserved, e.g. wave function parity or (crystal) momentum.

An important difference between discrete and continuous symmetries is that the latter can be built from infinitesimal transformations, i.e., they form a 'Lie group'. The infinitesimal transformations (2.3) in turn give us access to a Hermitian operator that corresponds to the conserved physical quantity. Examples for these so-called 'generators' of a transformation are momentum for translationally invariant systems and angular momentum for the case of rotational invariance.

Both for translations and for rotations, we found three independent generators, namely the x-, y- and z- components of $\hat{\vec{P}}$ and $\hat{\vec{L}}$. That general translations can be obtained from ones along x, y and z axes is immediately intuitive, because finite translations can easily be combined. That three components remain sufficient for general rotations is maybe less immediately clear, because combining finite rotations is not as straightforward. Nevertheless, real-space rotations can be described with three parameters, giving three independent generators, and this was found to carry over to wave-function rotations, see Sec. (2.2.1). The number of independent generators is sometimes called the dimension of a Lie group.

The mathematical structure describing the generators is called a 'Lie algebra', its most essential aspect for us are the commutation relations between generators. Translations and accordingly components of (linear) momentum commute, the resulting (Lie-)group is 'Abelian' and eigenstates common to all three momenta can be found. Rotations and angular-momentum components, on the other hand, do not commute but have characteristic commutation relations (2.39). Eigenstates common to all three components can thus not be expected, in the next chapter we are going to analyze this problem.

3 Angular momentum

Le Bellac, Sakurai I, Nolting (QM)

3.1 Revision of Spin

The spin is typically introduced by discussing the Stern-Gerlach (SG) experiment, where a beam of particles with spin 1/2 (silver) passes through a magnetic field and is consequently split. The aspects important to our discussion are

1. Results are quantized, the spin turns out to be $\pm\frac{\hbar}{2}$.
2. Sending the $+\hbar/2$ beam through a second SG apparatus with the same quantization axis gives only a signal with $+\hbar/2$, the $-\hbar/2$ signal is now missing.
3. If a second SG apparatus along x is put after the first (along z), it turns out that each beam (the one with $\langle S_z\rangle = +\hbar/2$ and the one with $-\hbar/2$ can again be split into $\langle S_x\rangle = \pm\hbar/2$.
4. If the first SG apparatus selects $\langle S_z\rangle = +\hbar/2$ and a second is used to select $\langle S_x\rangle = +\hbar/2$, then a third SG apparatus along z finds again *both* states $\langle S_z\rangle = \pm\hbar/2$ with equal frequency.

The most important (for our discussion) conclusions drawn from the experiments are

1. The spin is a vector: The axis $\vec{n}$ of a SG apparatus can be rotated in space and one always finds $\pm\hbar/2$ along this selected axis, i.e. results rotate like they would for a vector. One can thus denote the result as $\frac{\hbar}{2}\vec{n}$.
2. It is a two-level system:
   - One always find only two values: $\pm\hbar/2$
   - A second SG apparatus along an orthogonal direction (e.g. x after z as in the fourth point above) undoes the selection done with the first, i.e., one can *not* combine the information of two orthogonal SG experiments to obtain $2 \times 2 = 4$ states.

These two levels reside in a space different from the 3D position space spanned by $|\vec{r}\rangle$. For each position-space wave function $\psi(\vec{r}) = \langle\vec{r}|\psi\rangle$, we can additionally choose the spin state, i.e., the full wave function in the combined space is $|\psi\rangle \otimes |\sigma\rangle$, a tensor product of a first part expressing $\psi(\vec{r})$ (but potentially in some other basis than $\vec{r}$) and a second part expressing the spin. We discuss here for the moment only the second part, i.e., all operators only act on this part of the wave function. Our goal is now to find the operators describing the spin along various axes in this two-level Hilbert space, where we make use of these two conclusions (vector nature and 2-level system).

3.1.1 Hermitian 2×2 matrices

We are looking for operators describing the spin. We would like to have a matrix for any axis orientation, as we are dealing with a 2-level system, these are Hermitian 2×2 matrices:

$$S = \begin{pmatrix} A & C \\ C^* & B \end{pmatrix} , \tag{3.1}$$

where Hermiticity requires A and B to be real.

The operators must moreover be traceless because the sum of their eigenvalues $\lambda_{1,2} = \pm\hbar/2$ vanishes. [1] We thus look for matrices of the form

$$\hat{S} = \frac{\hbar}{2} \begin{pmatrix} a & |b|\mathrm{e}^{-i\alpha} \\ |b|\mathrm{e}^{i\alpha} & -a \end{pmatrix} \tag{3.2}$$

with real numbers a, α and $|b| > 0$.

The characteristic polynomial of the matrix part is $\lambda^2 - a^2 - |b|^2 = 0$ and since $\lambda^2 = 1$, $a^2 + |b|^2 = 1$, suggesting the parametrization

$$\hat{S} = \frac{\hbar}{2} \begin{pmatrix} \cos\beta & \sin\beta\mathrm{e}^{-i\alpha} \\ \sin\beta\mathrm{e}^{i\alpha} & -\cos\beta \end{pmatrix} . \tag{3.3}$$

The eigenvalues of this matrix are by construction $\pm\frac{\hbar}{2}$, the eigenvectors can be written as

$$|v_1\rangle = \begin{pmatrix} \mathrm{e}^{-i\frac{\alpha}{2}} \cos\frac{\beta}{2} \\ \mathrm{e}^{i\frac{\alpha}{2}} \sin\frac{\beta}{2} \end{pmatrix} , \quad |v_2\rangle = \begin{pmatrix} -\mathrm{e}^{-i\frac{\alpha}{2}} \sin\frac{\beta}{2} \\ \mathrm{e}^{i\frac{\alpha}{2}} \cos\frac{\beta}{2} \end{pmatrix} . \tag{3.4}$$

Equation (3.3) describes a general spin-$\frac{1}{2}$ operator along some arbitrary direction, as we have not used any information that would tie it to x, y or z directions. Specific choices for the parameters $\alpha = \alpha_{x/y/z}$ and $\beta = \beta_{x,y,z}$, which we have to determine, will give the matrices determining the spin along x-, y- and z- axes.

We get one of the three matrices $S_{x/y/z}$ almost for free, because we can define our basis in the 2×2 vector space: As the second of two SG experiments with the same axis only sees the state selected by the first (see point 2 above), the first apparently has prepared the spins into an eigenstate, the two eigenvalues are $\pm\hbar/2$. The corresponding eigenstates can be called "up" and "down", $|\uparrow\rangle$ and $|\downarrow\rangle$ or $|+\rangle$ and $|-\rangle$. In this basis, the operator describing a SG apparatus (along this selected axis, which we will call z), can thus be written as

$$\hat{S}_z = \frac{\hbar}{2}|\uparrow\rangle\langle\uparrow| + (-\frac{\hbar}{2})|\downarrow\rangle\langle\downarrow| = \frac{\hbar}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} . \tag{3.5}$$

This amounts to choosing to work in the eigenbasis of the $\hat{S}_z$ operator. Comparison to Eq. (3.3) suggests $\beta_z = 0$ (α_z undefined and unnecessary) in this basis; $\beta = \pi$ would be an equally valid choice. We still need matrices $S_{x/y}$ expressed in the eigenbasis of S_z.

[1] The trace is clearly given by $\lambda_1 + \lambda_2$ in the eigenbasis; as traces are invariant under basis transformations, it also vanishes for an arbitrary basis.

3.1.2 Vector-nature of the spin

If we rotate the SG apparatus away from z and point it along axis $\vec{n}$ (with $|\vec{n}| = 1$), we expect to measure $\pm\hbar/2$ along this axis and the experimental result $\vec{S} = \pm\hbar/2\vec{n}$ can be projected onto a coordinate system by using

$$\vec{n} = \begin{pmatrix} \sin\theta\cos\phi \\ \sin\theta\sin\phi \\ \cos\theta \end{pmatrix} . \tag{3.6}$$

For a beam polarized into the $|\vec{n}\uparrow\rangle$ state, the expectation values of the spin along x, y and z (instead of along $\vec{n}$) are then the projections of $\frac{\hbar}{2}\vec{n}$ onto the x-, y, and z-axes and given by

$$\begin{aligned} \langle\vec{S}\rangle = \frac{\hbar}{2}\vec{n} \quad \Rightarrow \quad \langle\vec{S}\rangle_x &= \frac{\hbar}{2}n_x = \frac{\hbar}{2}\sin\theta\cos\phi, \\ \langle\vec{S}\rangle_y &= \frac{\hbar}{2}n_y = \frac{\hbar}{2}\sin\theta\sin\phi \\ \text{and} \quad \langle\vec{S}\rangle_z &= \frac{\hbar}{2}n_z = \frac{\hbar}{2}\cos\theta\ . \end{aligned} \tag{3.7}$$

(Analogously for a beam prepared in $|\vec{n}\downarrow\rangle$.)

On the other hand, we could calculate the expectation values from the eigenstate Eq. (3.4), denoted by $|\vec{n},\uparrow\rangle = |v_1\rangle$ (and analogously from $|\vec{n},\downarrow\rangle = |v_2\rangle$ for the other spin projection). The expectation value of $\hat{S}_z$ is then obtained by using Eq. (3.5) and gives:

$$\begin{aligned} \langle\vec{n}\uparrow|\hat{S}_z|\vec{n}\uparrow\rangle &= \frac{\hbar}{2}(\mathrm{e}^{-i\frac{\alpha}{2}}\cos\frac{\beta}{2}, \mathrm{e}^{i\frac{\alpha}{2}}\sin\frac{\beta}{2})^* \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} \mathrm{e}^{-i\frac{\alpha}{2}}\cos\frac{\beta}{2} \\ \mathrm{e}^{i\frac{\alpha}{2}}\sin\frac{\beta}{2} \end{pmatrix} = \\ &= \frac{\hbar}{2}(\mathrm{e}^{-i\frac{\alpha}{2}}\cos\frac{\beta}{2}, \mathrm{e}^{i\frac{\alpha}{2}}\sin\frac{\beta}{2})^* \begin{pmatrix} \mathrm{e}^{-i\frac{\alpha}{2}}\cos\frac{\beta}{2} \\ -\mathrm{e}^{i\frac{\alpha}{2}}\sin\frac{\beta}{2} \end{pmatrix} = \frac{\hbar}{2}\left(\cos^2\frac{\beta}{2} - \sin^2\frac{\beta}{2}\right) = \\ &= \frac{\hbar}{2}\cos\beta \end{aligned} \tag{3.8}$$

If the spin is a vector, these two procedures have to give the same, i.e., $\langle S\rangle_z = \langle S_z\rangle$. Comparison to (3.7) yields $\beta = \theta$ for the spin along an arbitrary axis $\vec{n}$; $\beta = -\theta$ would be an equally valid choice. $\beta_z = 0$, see above, would correspond to $\vec{n}$ along the z axis.

In the case of $\hat{S}_x$ and $\hat{S}_y$, we do not know the matrices whose expectation values we want, but we do know that $\vec{n}$ would have to be turned into the plane orthogonal to z to polarize a beam along either direction, Consequently, $\theta = \pi/2$ in Eqs. (3.6) and (3.7), with the accordingly chosen $\beta_{x,y} = \pi/2$ in (3.3) giving operators of the form

$$\hat{S}_x = \frac{\hbar}{2}\begin{pmatrix} 0 & \mathrm{e}^{-i\alpha_x} \\ \mathrm{e}^{i\alpha_x} & 0 \end{pmatrix} \quad \text{and} \quad \hat{S}_y = \frac{\hbar}{2}\begin{pmatrix} 0 & \mathrm{e}^{-i\alpha_y} \\ \mathrm{e}^{i\alpha_y} & 0 \end{pmatrix} . \tag{3.9}$$

We use now these operators to calculate expectation values for a beam along an axis $\vec{n}$ lying the x-y plane. [2] Using $\beta = \pi/2$, we find

$$\langle\vec{n}\uparrow|\hat{S}_{x,y}|\vec{n}\uparrow\rangle = \frac{\hbar}{2\sqrt{2}^2}(\mathrm{e}^{-i\frac{\alpha}{2}}, \mathrm{e}^{i\frac{\alpha}{2}})^* \begin{pmatrix} 0 & \mathrm{e}^{-i\alpha_{x,y}} \\ \mathrm{e}^{i\alpha_{x,y}} & 0 \end{pmatrix} \begin{pmatrix} \mathrm{e}^{-i\frac{\alpha}{2}} \\ \mathrm{e}^{i\frac{\alpha}{2}} \end{pmatrix} =$$

[2] This restriction is not necessary, but simplifies the calculation without spoiling the argument.

$$= \frac{\hbar}{4}(\mathrm{e}^{-i\frac{\alpha}{2}}, \mathrm{e}^{i\frac{\alpha}{2}})^* \begin{pmatrix} \mathrm{e}^{-i\alpha_{x,y}}\mathrm{e}^{i\frac{\alpha}{2}} \\ \mathrm{e}^{i\alpha_{x,y}}\mathrm{e}^{-i\frac{\alpha}{2}} \end{pmatrix} = \frac{\hbar}{4}\left(\mathrm{e}^{i(\frac{\alpha}{2}-\alpha_{x,y}+\frac{\alpha}{2})} + \mathrm{e}^{i(-\frac{\alpha}{2}+\alpha_{x,y}-\frac{\alpha}{2})}\right) =$$

$$= \frac{\hbar}{2}\cos(\alpha - \alpha_{x,y}) \tag{3.10}$$

and comparison to (3.7) yields $\alpha_x = 0$, $\alpha_y = \pi/2$ as a consistent choice, together with $\alpha = \phi$: the expectation values become $\langle \hat{S}_x \rangle = \cos\phi$ and $\langle \hat{S}_y \rangle = \cos(\phi - \pi/2) = \sin\phi$. Using a beam polarized along a general direction $\vec{n}$, one can easily check that matrices (3.3) with $\beta_z = 0, \beta_{x,y} = \pi/2$, $\alpha_x = 0$, $\alpha_y = \pi/2$ leads to measurements consistent with (3.7). These parameters describe thus a *valid* set of spin matrices, however, not a unique one, see Sec. 3.1.4.

We have found the spin to be described by three matrices

$$\hat{S}_x = \frac{\hbar}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \frac{\hbar}{2}\sigma_x = \frac{\hbar}{2}\sigma_1 \tag{3.11}$$

$$\hat{S}_y = \frac{\hbar}{2}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} = \frac{\hbar}{2}\sigma_y = \frac{\hbar}{2}\sigma_2 \tag{3.12}$$

$$\hat{S}_z = \frac{\hbar}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \frac{\hbar}{2}\sigma_z = \frac{\hbar}{2}\sigma_3 \,. \tag{3.13}$$

The matrices $\sigma_{x,y,z}$, often referred ot as $\sigma_{1,2,3}$ are the Pauli matrices, they span the space of traceless Hermitian 2×2 matrices; they are said to "generate the group SU(2)". Together with the 2×2 unit matrix, sometimes referred to as σ_0, they can be used to express any 2×2 Hermitian matrix.

Different spin components do not commute. Instead, their commutation relations are given by

$$[\hat{S}_k, \hat{S}_l] = i\hbar\epsilon_{klm}\hat{S}_m \,, \quad \text{resp.} \quad [\sigma_k, \sigma_l] = 2i\epsilon_{klm}\sigma_m \,. \tag{3.14}$$

A second important property is

$$\sigma_\gamma^2 = \mathbb{I} \,. \tag{3.15}$$

Fulfilling these properties is the actual criterion for a valid set of spin matrices. As the commutation relations for $\hat{\vec{S}}$ turn out to be same as those for (orbital) angular momentum $\hat{\vec{S}}$, see Eq. (2.39), the group SU(2) should have very similar properties as the group SO(3) of rotations. This is explored in the following subsection.

3.1.3 Spin and Rotations

The sought-after rotations should be written in an analogous way to rotations of vectors $\vec{r}$ or of wave-functions in position space, see Eq. (2.35), as the Lie-algebras are equivalent. The new rotations written using spin-$\frac{1}{2}$ matrices, which act on the spin part of the wave function, are

$$R_{\vec{n}}(\alpha) = \mathrm{e}^{-\frac{i}{\hbar}\alpha\vec{n}\hat{\vec{S}}} = \mathrm{e}^{-i\frac{\alpha}{2}\vec{n}\vec{\sigma}} \,, \tag{3.16}$$

where $\vec{\sigma}$ is the vector of Pauli matrices. Before writing a series expansion for this, let us note that the relation for the squared Pauli matrices, Eq. (3.15), carries over to $\vec{n}\vec{\sigma}$:

$$(\vec{n}\vec{\sigma})^2 = n_x^2\sigma_x^2 + n_y^2\sigma_y^2 + n_z^2\sigma_z^2 = (n_x^2 + n_y^2 + n_z^2)\mathbb{I} = \mathbb{I}\,. \tag{3.17}$$

$(\vec{n}\vec{\sigma})^m$ is thus $\mathbb{I}$ for even m and $\vec{n}\vec{\sigma}$ for odd m. The rotation then becomes

$$\begin{aligned} R_{\vec{n}}(\alpha) &= \sum_{m=0}^{\infty} \frac{(-i\frac{\alpha}{2})^m}{m!}(\vec{n}\vec{\sigma})^m = \mathbb{I}\sum_{m=0}^{\infty}\frac{(-i\frac{\alpha}{2})^{2m}}{(2m)!} + \vec{n}\vec{\sigma}\sum_{m=0}^{\infty}\frac{(-i\frac{\alpha}{2})^{2m+1}}{(2m+1)!} = \\ &= \mathbb{I}\underbrace{\sum_{m=0}^{\infty}\frac{(-1)^m}{(2m)!}\left(\frac{\alpha}{2}\right)^{2m}}_{\cos\frac{\alpha}{2}} - i\vec{n}\vec{\sigma}\underbrace{\sum_{m=0}^{\infty}\frac{(-1)^m}{(2m+1)!}\left(\frac{\alpha}{2}\right)^{2m+1}}_{\sin\frac{\alpha}{2}} = \\ &= \cos\frac{\alpha}{2}\,\mathbb{I} - i\sin\frac{\alpha}{2}\,\vec{n}\vec{\sigma} = \begin{pmatrix} \cos\frac{\alpha}{2} - in_z\sin\frac{\alpha}{2} & (-in_x - n_y)\sin\frac{\alpha}{2} \\ (-in_x + n_y)\sin\frac{\alpha}{2} & \cos\frac{\alpha}{2} + in_z\sin\frac{\alpha}{2} \end{pmatrix}, \end{aligned} \tag{3.18}$$

where the relation to $\sin\frac{\alpha}{2}$ and $\cos\frac{\alpha}{2}$ are found either from the serie expansions of these functions or from comparing $\cos\frac{\alpha}{2}$ and $\sin\frac{\alpha}{2}$ to the real and imaginary parts of the series for $\mathrm{e}^{i\alpha/2}$.

To see whether $R_{\vec{n}}(\alpha)$ indeed describes - as we expect - rotations of a spin, let's calculate a few examples of rotations around x and apply them to the state $|\uparrow\rangle$:

$$R_x(\alpha)|\uparrow\rangle = \begin{pmatrix} \cos\frac{\alpha}{2} & -i\sin\frac{\alpha}{2} \\ -i\sin\frac{\alpha}{2} & \cos\frac{\alpha}{2} \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \cos\frac{\alpha}{2} \\ -i\sin\frac{\alpha}{2} \end{pmatrix}. \tag{3.19}$$

With a rotation angle of $\alpha = \pi/2$, the state indeed turns out to be in the x-y plane and in fact parallel to y, because it is the eigenstate of σ_y corresponding to $\lambda = 1$:

$$\sigma_y R_x(\pi/2)|\uparrow\rangle = \begin{pmatrix} 0 & -i \\ -i & 0 \end{pmatrix}\begin{pmatrix} 1/\sqrt{2} \\ -i/\sqrt{2} \end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -i \end{pmatrix}. \tag{3.20}$$

Rotating $|\uparrow\rangle$ further by π gets us to

$$R_x(\pi)|\uparrow\rangle = \begin{pmatrix} 0 \\ -i \end{pmatrix} = i|\downarrow\rangle\,, \tag{3.21}$$

which is equivalent to state $|\downarrow\rangle$, as quantum mechanical states may always be multiplied by a phase. Finally, rotation by 2π leads to $(\cos\pi, \sin\pi) = -|\uparrow\rangle$, which is similarly equivalent to the original $|\uparrow\rangle$. That rotation of a state along $+z$ by angles $\pi/2$, π, and 2π leads to states along y, $-z$ and z is indeed exactly what we expect for a rotation around the x-axis. That rotating by 2π does not lead to exactly the same state as the one we start with – but adds a phase -1 – is admittedly a bit strange, but quantum mechanical wave functions do not have to be unique, but are only defined up to an arbitrary phase.

Here, orientation in 'spin space' is only related to orientation in position space insofar as we have decided to call "up" a state polarized along $+z$ direction. If we change that definition, everything else carries over in the same manner. In particular, the direction chosen for "up" does here not have to be the "+z" direction chosen when describing the particle's wave function. We are going to see in Sec. 9.5.1 that rotations in spin and in position space are actually connected to each other if one starts from the Dirac equation.

3.1.4 Alternatives

Alternative choices of α_x, α_y and α are clearly possible in Eq.(3.10), they correspond to coordinate systems where the x- and y-axes of the lab are rotated. These alternatives are equally valid, as the spin does not know what x- and y-axes we might choose. Additionally, further choices for all three spin components are possible by allowing the z-axis, which we chose as the axis where $|\uparrow\rangle = (1,0)$, to be something else apart form "straight up". (The spin does here not know about gravity.)

Any rotated coordinate system gives an equally valid set of spin matrices. The all-important commutation relations are not affected by such rotations of the coordinate system, because the rotations are described by unitary operators U. (In the previous section, we discussed as an example the unitary transformation (3.16) as a rotation around x.) For Pauli matrices acting in a rotated frame, $\tilde{\sigma}_i = U^\dagger \sigma_i U$, the commutation relations are

$$\begin{aligned}[\tilde{\sigma}_i, \tilde{\sigma}_j] &= U^\dagger \sigma_i \underbrace{UU^\dagger}_{=\mathbb{I}} \sigma_j U - U^\dagger \sigma_j U U^\dagger \sigma_i U = U^\dagger [\sigma_i, \sigma_j] U = U^\dagger 2i\epsilon_{i,j,k}\sigma_k U = \\ &= 2i\epsilon_{i,j,k}\tilde{\sigma}_k \end{aligned} \tag{3.22}$$

and also the square of the matrices remains $\mathbb{I}$

$$\tilde{\sigma}_i^2 = U^\dagger \sigma_i \underbrace{UU^\dagger}_{=\mathbb{I}} \sigma_i U = U^\dagger \underbrace{\sigma_i \sigma_i}_{=\mathbb{I}} U = U^\dagger U = \mathbb{I}\,. \tag{3.23}$$

One sees that the equivalence of "Pauli matrices" obtained from rotated coordinate systems is evident from the preserved algebraic properties.

3.2 General angular momentum

Angular-momentum operators redefined by three operators J_x, J_y and J_z that obey commutation relations like the generators of the group of rotations,

$$[J_k, J_l] = i\epsilon_{klm} J_m\,. \tag{3.24}$$

These are also the commutation relations for orbital angular momentum $\vec{L} = \vec{r} \times \vec{p}$ and spin $\vec{S}$, if they are given in multiples of $\hbar$:

$$\left[\frac{1}{\hbar}L_k, \frac{1}{\hbar}L_l\right] = \frac{1}{\hbar^2}[L_k, L_l] = \frac{1}{\hbar^2} i\hbar\epsilon_{klm} L_m = i\epsilon_{klm}\frac{1}{\hbar}L_m\,. \tag{3.25}$$

For the rest of this chapter, we are going to this system of units, which corresponds to setting $\hbar = 1$, and it will only be written in exceptional cases.

As some Hamiltonians H are rotationally invariant (e.g. Coulomb potential), a reasonable strategy would be to first find the eigensystem of the generators of rotation and then diagonalize the Hamiltonian in this eigensystem. Ideally, we would like a common eigensystem of all operators known to commute with H, i.e., of J_x, J_v and J_z. However, as the components of $\vec{J}$ do not commute, they cannot be diagonalized together and we cannot hope to achieve this. We can try to get as close as possible to this goal, which means finding a basis, where

the three conserve quantities are "as diagonal as possible", which is then called "irreducible representation".

To do so, we start out in the eigenbasis of one of the components, usually J_z; the z-axis is then called the "quantization axis" of angular momentum. Additionally, we can diagonalize any operator that commutes with J_z, and in order not to spoil relations with J_x and J_y, we consider

$$J^2 = J_x^2 + J_y^2 + J_z^2 \tag{3.26}$$

that commutes with all of them. We can for example check J_x:

$$\begin{aligned}[J^2, J_x] = [J_x^2 + J_y^2 + J_z^2, J_x] = 0 + J_y[J_y, J_x] + [J_y, J_x]J_y + J_z[J_z, J_x] + [J_z, J_x]J_z = \\ = -iJ_yJ_z - iJ_zJ_y + iJ_zJ_y + iJ_yJ_z = 0\ .\end{aligned} \tag{3.27}$$

J^2 It is called a 'scalar', see Sec. 3.3, because it commutes with any rotation. It thus nicely expresses rotational invariance, which motivates our choice to investigate it.

Eigenvalues of J^2 cannot be negative, as they can be written as a sum of norms

$$\begin{aligned}C_{J^2}\langle\psi_{J^2}|\psi_{J^2}\rangle = \langle\psi_{J^2}|J^2|\psi_{J^2}\rangle = \langle\psi_{J^2}|J_x^2|\psi_{J^2}\rangle + \langle\psi_{J^2}|J_y^2|\psi_{J^2}\rangle + \langle\psi_{J^2}|J_z^2|\psi_{J^2}\rangle = \\ = ||J_x|\psi_{J^2}\rangle||^2 + ||J_y|\psi_{J^2}\rangle||^2 + ||J_z|\psi_{J^2}\rangle||^2 \geq 0\ .\end{aligned} \tag{3.28}$$

We denote the eigenvalues of J^2 by $j(j+1)$ (with $j \geq 0$), because this turns out to be convenient later and imposes no additional restriction. Eigenvalues of J_z are denoted by m and the corresponding common eigenstates by $|j, m\rangle$. As the next step, we try to find the allowed eigenvalues, where we make use of an approach analogous to that used for the harmonic oscillator.

In this basis, where J^2 and J_z can be replaced by their eigenvalues, Eq. (3.26) becomes an eigenvalue equation for the operator $J_x^2 + J_y^2$:

$$(J_x^2 + J_y^2)|j, m\rangle = (J^2 - J_z^2)|j, m\rangle = \left[j(j+1) - m^2\right]|j, m\rangle\ . \tag{3.29}$$

With its sum of two squared operators that cannot be diagonalized at the same time, this equation looks formally like the time-independent Schrödinger equation of the harmonic oscillator in the eigenbasis of the Hamiltonian $(\hat{x}^2 + \hat{p}^2)|n\rangle = \frac{2E_n}{\hbar\omega}|n\rangle$. In analogy to the procedure followed in Sec. 1.5.1, one also introduces similar operators

$$J_+ = J_x + iJ_y \quad \text{and} \tag{3.30}$$

$$J_- = J_x - iJ_y = J_+^\dagger \tag{3.31}$$

$$\tag{3.32}$$

here. Together with $J_0 = J_z$, they have commutator relations

$$[J_0, J_\pm] = \underbrace{[J_z, J_x]}_{iJ_y} \pm i\underbrace{[J_z, J_y]}_{-iJ_x} = \pm(J_x \pm iJ_y) = \pm J_\pm \quad \text{and} \tag{3.33}$$

$$[J_+, J_-] = i\underbrace{[J_y, J_x]}_{-iL_z} - i[J_x, J_y] = 2J_0\ , \tag{3.34}$$

and of course all commute with J^2, which can be written as

$$J^2 = J_z^2 + \frac{1}{4}(J_+ + J_-)^2 + \frac{1}{-4}(J_+ - J_-)^2 = J_0^2 + \frac{1}{2}(J_+J_- + J_-J_+) \tag{3.35}$$

In the basis $|j,m\rangle$ diagonalizing both J^2 and J_z, we then again find an equation

$$J_+J_-|j,m\rangle = J^2 - J_0(J_0-1)|j,m\rangle = [j(j+1) - m(m-1)]\,|j,m\rangle \tag{3.36}$$

rather similar to its harmonic-oscillator counter part. The similarity can be seen by comparison with the eigenvalue equation $H|n\rangle = \hbar\omega(a^\dagger a + \frac{1}{2}) = E|n\rangle$, and consequently $a^\dagger a|n\rangle = [E/(\hbar\omega)) - 1/2]|n\rangle$, which is obtained from Eqs. (1.50) and (1.51).

As in the case of the harmonic oscillator, applying the operators $J_\pm$ to an eigenstate $|j,m\rangle$ leads to other eigenstates. Here,

$$J_0J_\pm|j,m\rangle = (J_\pm J_0 \pm J_\pm)|j,m\rangle = J_\pm(J_0 \pm \mathbb{I})|j,m\rangle = J_\pm(m\pm 1)|j,m\rangle = (m\pm 1)J_\pm|j,m\rangle\ , \tag{3.37}$$

i.e. one obtains an eigenstate of J_0, where the eigenvalue m is raised or lowered by 1. As $J_\pm$ and J^2 commute, this new state is still an eigenstate of J^2 and the eigenvalue $j(j+1)$ is unchanged.

In the case of the harmonic oscillator, we found the lowest eigenstate by noting that the norm of all vectors needs to be positive, and this is also important here. Assuming that we have one eigenstate $|j,m\rangle$ and that it is normalized $|||j,m\rangle|| = 1$, we can continue from there.

$$\begin{aligned} ||J_+|j,m\rangle||^2 &= \langle j,m|J_+^\dagger J_+|j,m\rangle = \langle j,m|J_-J_+|j,m\rangle = \langle j,m|\left(J^2 - J_0(J_0+1)\right)|j,m\rangle = \\ &= [j(j+1) - m(m+1)]\underbrace{\langle j,m|j,m\rangle}_{=1} \geq 0 \end{aligned} \tag{3.38}$$

In order for this to be positive and taking into account that $j \geq 0$, m must be $\leq j$. On the other hand,

$$\begin{aligned} ||J_-|j,m\rangle||^2 &= \langle j,m|J_-^\dagger J_-|j,m\rangle = \langle j,m|J_+J_-|j,m\rangle = \langle j,m|\left(J^2 - J_0(J_0-1)\right)|j,m\rangle = \\ &= j(j+1) - m(m-1) \geq 0 \end{aligned} \tag{3.39}$$

has to hold, which requires $m \geq -j$. In contrast to the harmonic oscillator, there is here a highest eigenstate ($m = j$) in addition to a lowest ($m = -j$) and the roles of J_- and J_+ are perfectly symmetric. Again as with the harmonic oscillator, "illegal" eigenstates with $|m| > j$ are prevented by requiring that $J_\pm|j,m\rangle = 0$ for any m that would give $|m\pm 1| > j$. This is automatically fulfilled if - and only if - m is restricted to values $-j$, $-j+1$, …, $j-1$, j. Applying J_+ (J_-) to $|j,j\rangle$ ($|j,-j\rangle$) then gives a vanishing result, see Eqs. (3.38) and (3.39) and the sequence terminates before reaching negative norms. As there has to be an integer number of steps between $m = -j$ and $m = +j$, $2j$ has to be integer itself. The quantum number j can be integer or half-integer.

Finally, Eqs. (3.38) and (3.39) also give us the norm of the eigenstates:

$$||J_\pm|j,m\rangle||^2 = j(j+1) - m(m\pm 1) \quad \Rightarrow \quad |j,m\pm 1\rangle = \frac{1}{\sqrt{j(j+1) - m(m\pm 1)}}J_\pm|j,m\rangle \tag{3.40}$$

3.2.1 Some properties of the eigenbasis

The angular-momentum operators can now be written in the eigenbasis of J^2 and J_z, i.e., in terms of the states $|j,m\rangle$. For a given j, they are $(2j+1)\times(2j+1)$ matrices of the form

$$\langle j',m'|J^2|j,m\rangle = j(j+1)\delta_{m,m'}\delta_{j,j'}\ , \tag{3.41}$$

$$\langle j',m'|J_0|j,m\rangle = m\ \delta_{m,m'}\delta_{j,j'} \quad \text{and} \tag{3.42}$$

$$\langle j',m'|J_\pm|j,m\rangle = \sqrt{j(j+1)-m(m\pm1)}\delta_{m,m'\mp1}\delta_{j,j'}\ . \tag{3.43}$$

Operators J_x and J_y can be obtained from $J_\pm$.

As an example, we look at the operators for $j=1/2$, which has possible $m\pm1/2$. $J_0=J_z$ is diagonal and can take values $\pm1/2$. The ladder operator

$$\langle\frac{1}{2},m'|J_+|\frac{1}{2},m\rangle = \sqrt{\frac{1}{2}\frac{3}{2}-m(m+1)}\delta_{m,m'-1} \tag{3.44}$$

can only be non-zero for $m=-1/2$ and $m'=1/2$, with $\langle\frac{1}{2},+\frac{1}{2}|J_+|\frac{1}{2},-\frac{1}{2}\rangle = \sqrt{\frac{3}{4}+\frac{1}{2}\frac{1}{2}}=1$. Analogously, $\langle\frac{1}{2},-\frac{1}{2}|J_-|\frac{1}{2},+\frac{1}{2}\rangle=1$ is the only non-zero element of $J_-=J_+^\dagger$. These can be combined into

$$J_x=\frac{1}{2}(J_++J_-)=\frac{1}{2}\begin{pmatrix}0&1\\1&0\end{pmatrix} \quad\text{and}\quad J_y=\frac{1}{2i}(J_+-J_-)=\frac{-i}{2}\begin{pmatrix}0&1\\-1&0\end{pmatrix}=\frac{1}{2}\begin{pmatrix}0&-i\\i&0\end{pmatrix}\ . \tag{3.45}$$

$J_{x,y,z}$ are thus Pauli matrices Eq. (3.11), the operators describing a spin 1/2.

Angular momentum is supposed to generate rotations, in particular $J_0=J_z$ around the z-axis. $|j,m\rangle$ is an eigenstate and the rotation is thus just a phase $R_z(\alpha)|j,m\rangle = \mathrm{e}^{-i\alpha J_z}|j,m\rangle = \mathrm{e}^{-i\alpha m}|j,m\rangle$. For the case of a rotation by 2π, this phase can either be 1 (integer m and thus j) or -1 (half-integer j). We had already seen in Sec. 3.1.3 that rotation by 2π of a spin-1/2 particle gives a phase -1, in agreement with $j=1/2$. For orbital angular momentum $L=r\times p$, we can find the phase by pushing the rotation onto the argument $\vec{r}$ of the wave function, see Eq. (2.22). One the finds

$$R_z(2\pi)\psi(\vec{r}) = \langle\vec{r}|R_z(2\pi)\psi\rangle = \psi(R_z^{-1}(2\pi)\vec{r}) = \psi(R_z(-2\pi)\vec{r}) = \psi(\vec{r}) \tag{3.46}$$

and as the phase is $+1$, orbital angular momentum has to correspond to integer j.

The eigenfunction of orbital angular momentum can be obtained from differential equations that are in turn found by considering $L_{x,y,z}$ as generators of rotation. They are called spherical harmonics and discussed in standard textbooks.

3.3 Scalars and Vectors

States with different j quantum number were never mixed by components of J, see (3.42) and (3.43). The reason is of course that the corresponding operator J^2 commutes with all J_k. As a consequence, it is also invariant under rotations, which are generated by J_k. Such a quantity is called 'scalar'. This can be generalized: an operator A is called scalar, if it is invariant under rotations:

$$\langle\phi'|A|\psi'\rangle = \langle\phi|U^\dagger AU|\psi\rangle \quad\Rightarrow\quad A=U^\dagger AU\ , \tag{3.47}$$

where $U = U(\alpha, \vec{n})$ is the unitary transformation describing the rotation. Focusing on infinitesimal rotations (2.3) generated by J_k, we find

$$A = (\mathbb{I} + \frac{i\epsilon}{\hbar} J_k) A (\mathbb{I} - \frac{i\epsilon}{\hbar} J_k) = A + \frac{i\epsilon}{\hbar}[J_k, A] + \mathcal{O}(\epsilon^2) \quad \Rightarrow \quad [J_k, A] = 0 \tag{3.48}$$

Similarly, we can now state more clearly what a vector is. It is given by three operators whose expectaion values transform like components of a vecor. Going to a matrix notation, the vector rotation $\mathcal{R} = \mathcal{R}(\alpha, \vec{n})$ is given by

$$\vec{r}' = \mathcal{R}(\alpha, \vec{n})\vec{r} \quad \text{resp.} \quad r_i' = \sum_{j=1}^{3} \mathcal{R}_{i,j} r_j \tag{3.49}$$

and the expectation vlaues are then likewise expected to fulfill

$$\langle \vec{V}' \rangle = \mathcal{R}(\alpha, \vec{n}) \langle \vec{V} \rangle \quad \text{resp.} \quad \langle \phi' | V_i | \psi' \rangle = \langle \phi | U^\dagger V_i U | \psi \rangle = \sum_{j=1}^{3} \mathcal{R}_{i,j} \langle \phi | V_j | \psi \rangle \ . \tag{3.50}$$

If we once more consider infinitesimal rotations around the z axis, the 'space' and wave-function rotations come from (2.29) and (2.24) as

$$\mathcal{R}_z(\epsilon) = \mathbb{I} - i\epsilon T_z \quad \text{and} \quad U_z(\epsilon) = \mathbb{I} - \frac{i\epsilon}{\hbar} L_z \ . \tag{3.51}$$

Remembering that spin worked similarly to orbital angular momentum, we generalize from $\vec{L}$ to $\vec{J}$ [3] and find

$$\begin{aligned} (\mathbb{I} + \frac{i\epsilon}{\hbar} J_z) \vec{V} (\mathbb{I} - \frac{i\epsilon}{\hbar} J_z) &= (\mathbb{I} - i\epsilon T_z) \vec{V} \\ \vec{V} + \sum_k \frac{i\epsilon}{\hbar} [J_z, V_k] \vec{e}_k &= \vec{V} - i\epsilon T_z \vec{V} = \vec{V} + \epsilon V_y \vec{e}_x - \epsilon V_x \vec{e}_y \end{aligned} \tag{3.52}$$

Comparison of the components yields

$$[J_z, V_x] = i\hbar V_y, \quad [J_z, V_y] = -i\hbar V_x \quad [J_z, V_z] = 0 \ , \tag{3.53}$$

which can be summarized to

$$[J_i, V_j] = i\hbar \epsilon_{ijk} V_k \ . \tag{3.54}$$

These commutation relations define a valid vector of operators.

3.4 Adding angular momenta

Interest in angular momentum comes from its relation rotational symmetry: its components are conserved if the Hamiltonian is invariant under rotations. When two angular momenta interact with each other, rotational invariance for each is usually lost. If a spin is, e.g., somehow coupled to another spin, space no longer looks isotropic to this first spin, because the direction, in which the second spin points, is special (and vice versa). However, the total system consisting of both spins together should still be rotationally invariant. Consequently, total angular momentum should be the more instructive quantity to look at.

[3] For a particle with both orbital angular momentum and spin, one can show in a tensor-product basis, see next section, that this generalization makes sense.

3.4.1 Two spins $\frac{1}{2}$

The total Hilbert space for two particles with spin 1/2 each consists of four states. A possible basis is to give the magnetic quantum number m for each spin m_a and m_b. Two more quantum numbers would be the eigenvalues of $J^2_{a,b}$, but as j_a and j_b remain constant here, they will be suppressed. The states are then tensor products $|m_a\rangle \otimes |m_b\rangle$ of the single-spin states $|\uparrow\rangle$ and $|\downarrow\rangle$:

$$\begin{aligned}
|\uparrow,\uparrow\rangle &= |m_a = \frac{1}{2}\rangle \otimes |m_b = \frac{1}{2}\rangle \\
|\uparrow,\downarrow\rangle &= |m_a = \frac{1}{2}\rangle \otimes |m_b = \frac{-1}{2}\rangle \\
|\downarrow,\uparrow\rangle &= |m_a = \frac{-1}{2}\rangle \otimes |m_b = \frac{1}{2}\rangle \\
|\downarrow,\downarrow\rangle &= |m_a = \frac{-1}{2}\rangle \otimes |m_b = \frac{-1}{2}\rangle
\end{aligned} \tag{3.55}$$

The rotation expressed by total spin $\vec{S} = \vec{S}_a + \vec{S}_b$ can indeed be used to describe the rotation of such a tensor-product state $|\phi\rangle = |\phi_a\rangle \otimes |\phi_b\rangle$. One has here to take care that operator $\hat{\vec{S}}_a$ only acts on the first spin and $\hat{\vec{S}}_b$ only on the second, the complete notation in terms of tensor products is $\hat{\vec{S}} = \hat{\vec{S}}_a \otimes \mathbb{I} + \mathbb{I} \otimes \hat{\vec{S}}_b$. Rotating both individually by the same angle α around the same axis $\vec{n}$ is expressed as

$$\left(\mathrm{e}^{-i\frac{\alpha}{\hbar}\vec{S}_a\vec{n}}|\phi_a\rangle\right) \otimes \left(\mathrm{e}^{-i\frac{\alpha}{\hbar}\vec{S}_b\vec{n}}|\phi_b\rangle\right) = \left(\mathrm{e}^{-i\frac{\alpha}{\hbar}\vec{S}_a\vec{n}} \otimes \mathrm{e}^{-i\frac{\alpha}{\hbar}\vec{S}_b\vec{n}}\right)|\phi_a\rangle \otimes |\phi_b\rangle \tag{3.56}$$

and as operators relating to different spin commute, this can be written as

$$\mathrm{e}^{-i\frac{\alpha}{\hbar}(\vec{S}_a+\vec{S}_b)\vec{n}}|\phi\rangle = \mathrm{e}^{-i\frac{\alpha}{\hbar}\vec{S}\vec{n}}|\phi\rangle, \tag{3.57}$$

i.e., $\hat{\vec{S}}$ generates the rotation of the total two-spin state $|\phi\rangle$.

As the total spin may be expected to be a relevant characterization of the two-spin system, one would like to find a different basis in terms of the total spin and one of its components, i.e., in terms of $(S_a + S_b)^2 = S^2$ and S_z. These respective quantum numbers will be called J and M here and should replace m_a and m_b. As the total z component is the sum of the z components of the two spins $S_z = S_{a,z} + S_{b,z}$, the states of (3.55) already are eigenstates of M as well. The largest and smallest $M = m_a + m_b$ are here $M = \pm 1$ and as $M \leq |J|$, the largest possible $J = 1$. On the other hand, we must have $J = 1$ states, because with only $J = 0$, we would not find $M = \pm 1$.

We thus expect that the $|\uparrow,\uparrow\rangle$ state is the $M = 1$ state belonging to $J = 1$ and analyze its properties in more detail by applying the angular-momentum operators.

$$S_z|\uparrow,\uparrow\rangle = (S_{a,z} + S_{b,z})|\uparrow\rangle \otimes |\uparrow\rangle = 1|\uparrow,\uparrow\rangle \tag{3.58}$$

$$S_+|\uparrow,\uparrow\rangle = (S_{a,+} + S_{b,+})|\uparrow\rangle \otimes |\uparrow\rangle = 0 \tag{3.59}$$

$$S_-|\uparrow,\uparrow\rangle = (S_{a,-} + S_{b,-})|\uparrow\rangle \otimes |\uparrow\rangle = \sqrt{1}|\downarrow,\uparrow\rangle + \sqrt{1}|\uparrow,\downarrow\rangle \tag{3.60}$$

and analogously

$$S_z|\downarrow,\downarrow\rangle = (S_{a,z} + S_{b,z})|\uparrow\rangle \otimes |\uparrow\rangle = -1|\downarrow,\downarrow\rangle \tag{3.61}$$

$$S_+|\downarrow,\downarrow\rangle = (S_{a,+} + S_{b,+})|\uparrow\rangle \otimes |\uparrow\rangle = |\uparrow,\downarrow\rangle + |\downarrow,\uparrow\rangle \tag{3.62}$$
$$S_-|\downarrow,\downarrow\rangle = (S_{a,-} + S_{b,-})|\uparrow\rangle \otimes |\uparrow\rangle = 0\,. \tag{3.63}$$

Finally, the (unnormalized) "middle" state arising here has $S_z(|\uparrow,\downarrow\rangle + |\downarrow,\uparrow\rangle) = 0$. This indeed looks similar to the behavior expected for three states with $J = 1$ and $M = \pm 1$. If we identify $|\uparrow,\uparrow\rangle$ with $|J = 1, M = 1\rangle$ and $|\downarrow,\downarrow\rangle$ with $|J = 1, M = -1\rangle$, the third state would be obtained by use of the ladder operators

$$J_-|J = 1, M = 1\rangle = \sqrt{1 \cdot 2 - 1 \cdot 0}\,|J = 1, M = 0\rangle \quad \text{or}$$
$$J_+|J = 1, M = -1\rangle = \sqrt{1 \cdot 2 - (-1) \cdot 0}\,|J = 1, M = 0\rangle \tag{3.64}$$

Comparison to Eqs. (3.60) and (3.62) implies $|J = 1, M = 0\rangle = \frac{1}{\sqrt{2}}(|\uparrow,\downarrow\rangle + |\downarrow,\uparrow\rangle)$ (which is normalized to 1) and the three states have the properties of a $J = 1$ system.

From the tensor-product basis, we find a fourth linearly independent basis state as $\frac{1}{\sqrt{2}}(|\uparrow,\downarrow\rangle - |\downarrow,\uparrow\rangle)$. Its properties with respect to angular-momentum operators are

$$S_z \frac{1}{\sqrt{2}}(|\uparrow,\downarrow\rangle - |\downarrow,\uparrow\rangle) = 0 \tag{3.65}$$

and consequently $M = 0$, as well as

$$S_+ \frac{1}{\sqrt{2}}(|\uparrow,\downarrow\rangle - |\downarrow,\uparrow\rangle) = \frac{1}{\sqrt{2}}\,(0 + |\uparrow,\uparrow\rangle - |\uparrow,\uparrow\rangle + 0) = 0 \tag{3.66}$$
$$S_- \frac{1}{\sqrt{2}}(|\uparrow,\downarrow\rangle - |\downarrow,\uparrow\rangle) = \frac{1}{\sqrt{2}}\,(|\downarrow,\downarrow\rangle + 0 - 0 - |\downarrow,\downarrow\rangle) = 0\,. \tag{3.67}$$

The expectation value of S^2, which can be obtained via Eq. (3.35), also vanishes, because J_0 as well as $J_\pm$ all give 0, giving $J(J+1) = 0$ and $J = 0$.

Summarizing, the new basis in terms of J and M is composed of one $J = 0$ state and three $J = 1$ states:

$$|J = 0, M = 0\rangle = \frac{1}{\sqrt{2}}(|\uparrow,\downarrow\rangle - |\downarrow,\uparrow\rangle) \qquad \text{singlet} \tag{3.68}$$
$$\left.\begin{aligned} &|J = 1, M = 1\rangle = |\uparrow,\uparrow\rangle \\ &|J = 1, M = 0\rangle = \tfrac{1}{\sqrt{2}}\,(|\uparrow,\downarrow\rangle + |\downarrow,\uparrow\rangle) \\ &|J = 1, M = -1\rangle = |\downarrow,\downarrow\rangle \end{aligned}\right\} \qquad \text{triplet} \tag{3.69}$$

In case of two spins, total angular momentum is usually also referred to as spin, i.e., the triplet has $S = 1$ and the singlet $S = 0$.

The two states with $M = 0$ cannot be written as tensor-product states, we will discuss this in more detail later. Their rotation is nevertheless determined by the total spin, as in (3.56), because the statement is valid for the full tensor-product basis and consequently also for linear combinations of tensor-product states.

3.4.1.1 A plausible interaction: Heisenberg exchange

An interaction of two spins that preserves total rotational symmetry of a two-spin system is $\hat{\vec{S}}_a \cdot \hat{\vec{S}}_b$, such an interaction is often called "symmetry allowed". One can easily see that

a Hamiltonian given by this interaction does not commute with the individual spins (i.e., rotational symmetry is not preserved for each spin) but does commute with the total spin $\hat{\vec{S}}_a + \hat{\vec{S}}_b$. For component x (and analogously for y and z), one gets the following commutators for the two spins

$$\begin{aligned}\left[\hat{S}_{a,x}, \hat{\vec{S}}_a \cdot \hat{\vec{S}}_b\right] &= \left[\hat{S}_{a,x}, \hat{S}_{a,y}\hat{S}_{b,y} + \hat{S}_{a,z}\hat{S}_{b,z}\right] = \left[\hat{S}_{a,x}, \hat{S}_{a,y}\right]\hat{S}_{b,y} + \left[\hat{S}_{a,x}, \hat{S}_{a,z}\right]\hat{S}_{b,z} = \\ &= i\hat{S}_{a,z}\hat{S}_{b,y} - i\hat{S}_{a,y}\hat{S}_{b,z} \qquad (3.70)\\ \left[\hat{S}_{b,x}, \hat{\vec{S}}_a \cdot \hat{\vec{S}}_b\right] &= \left[\hat{S}_{b,x}, \hat{S}_{a,y}\hat{S}_{b,y} + \hat{S}_{a,z}\hat{S}_{b,z}\right] = \hat{S}_{a,y}\left[\hat{S}_{b,x}, \hat{S}_{b,y}\right] + \hat{S}_{a,z}\left[\hat{S}_{b,x}, \hat{S}_{b,z}\right] = \\ &= i\hat{S}_{a,y}\hat{S}_{b,z} - i\hat{S}_{a,z}\hat{S}_{b,y}\ . \qquad (3.71)\end{aligned}$$

Each commutator is different from 0, but their sum, i.e., $[\hat{S}_{a,x} + \hat{S}_{b,x}, \hat{S}_{a,x}]$ vanishes, showing that this interaction preserves total angular momentum.

The interaction Hamiltonian, called "Heisenberg Hamiltonian", is diagonalized in the basis given by J and M in addition to $S_a = S_b = \frac{1}{2}$. Its eigenvalue can be obtained from $S^2 = J(J+1)$ and $\hat{\vec{S}}_a^2 = \frac{1}{2}\cdot\frac{3}{2}$ as

$$S^2 = \left(\hat{\vec{S}}_a + \hat{\vec{S}}_b\right)^2 = \hat{\vec{S}}_a^2 + \hat{\vec{S}}_b^2 + 2\hat{\vec{S}}_a \cdot \hat{\vec{S}}_b \qquad (3.72)$$

$$\hat{\vec{S}}_a \cdot \hat{\vec{S}}_b |J, M\rangle = \left(\frac{\hat{\vec{S}}^2}{2} - \hat{\vec{S}}_a^2\right)|J, M\rangle = \left(\frac{J(J+1)}{2} - \frac{3}{4}\right)|J, M\rangle = \begin{cases} -\frac{3}{4}|J=0, M=0\rangle \\ \frac{1}{4}|J=1, M\rangle \end{cases} \qquad (3.73)$$

3.4.2 Spin $\frac{1}{2}$ and orbital angular momentum 1

Even though orbital angular momentum acts on the position-space part of the wave function and spin on the spin space, the two kinds of angular momenta can be coupled due to relativistic effects, as we will discuss later. The "z" axis for the spin can then no longer be assigned completely arbitrarily, as the spin then "knows" about directions in position space. As in the case of two interacting spins, total rotational symmetry will, however, ofter be preserved and the dominant interaction $\hat{\vec{L}} \cdot \hat{\vec{S}}$ also has a similar form. As in the preceding subsection, one can show that it preserves the components of $\hat{\vec{J}}_{tot}$, but not those of $\hat{\vec{L}}$ and $\hat{\vec{S}}$ separately.

Again, we would thus be interested in going from the basis in terms of $S = \frac{1}{2}$, $L = 1$, m_s (the magnetic quantum number of the spin) and m_l to a new basis in terms of $S = \frac{1}{2}$, $L = 1$, $J_{\rm tot} = L + S$ and M (the magnetic quantum number of $J_{\rm tot}$). The tensor-product basis is here

$$\begin{array}{lll} |m_l = 1, m_s = \frac{1}{2}\rangle, & |m_l = 0, m_s = \frac{1}{2}\rangle, & |m_l = -1, m_s = \frac{1}{2}\rangle \\ |m_l = 1, m_s = -\frac{1}{2}\rangle, & |m_l = 0, m_s = -\frac{1}{2}\rangle, & |m_l = -1, m_s = -\frac{1}{2}\rangle\ . \end{array}$$

As before, it makes sense to start from the state with maximal (minimal) $M = m_s + M_l = \pm\frac{3}{2}$, as they determine the maximal M and $J_{\rm tot}$. The operators of total angular momentum then give

$$J_{{\rm tot},z}|1, \frac{1}{2}\rangle = (m_l + m_s)|1, \frac{1}{2}\rangle = \frac{3}{2}|1, \frac{1}{2}\rangle \qquad (3.74)$$

$$J_{+,z}|1,\frac{1}{2}\rangle = (L_+ + S_+)|1,\frac{1}{2}\rangle = 0 \tag{3.75}$$

$$J_{-,z}|1,\frac{1}{2}\rangle = (L_- + S_-)|1,\frac{1}{2}\rangle = \sqrt{2}\,|0,\frac{1}{2}\rangle + |1,-\frac{1}{2}\rangle\,, \tag{3.76}$$

where the last line should be compared to

$$J_-|J=\frac{3}{2},M=\frac{3}{2}\rangle = \underbrace{\sqrt{\frac{3}{2}\frac{5}{2}-\frac{3}{2}\frac{1}{2}}}_{=\sqrt{3}}\,|J=\frac{3}{2},M=\frac{1}{2}\rangle \tag{3.77}$$

to find $|J=\frac{3}{2},M=\frac{1}{2}\rangle = \frac{1}{\sqrt{3}}(\sqrt{2}\,|0,\frac{1}{2}\rangle + |1,-\frac{1}{2}\rangle)$.

The orthogonal linear combinations of states with $M=\frac{1}{2}$ cannot belong to a second $J=\frac{3}{2}$ sequence, because it would not have any $M=\frac{3}{2}$ partner. It can, instead, be taken as the positive-M state of a $J=\frac{1}{2}$ doublet. The analogous considerations starting from the $J=\frac{3}{2},M=\frac{3}{2}\rangle$ state would give its $M=-\frac{1}{2}$ partner.

The combination of $\hat{\vec{L}}=1$ and $\hat{\vec{S}}=\frac{1}{2}$ consequently gives for the total angular momentum the possible values $J=\frac{3}{2}$, $J=\frac{1}{2}$, and the corresponding M for each. The eigenstates except the ones with largest and smallest M involve coefficients that can be obtained by repeatedly applying $J_\pm$ and arise through the $\sqrt{J(J+1)-M(M\pm1)}$-prefactor.

3.4.3 General case

The general procedure of adding to angular momenta J_a and J_b follows the same lines: The largest possible $\hat{\vec{J}}=\hat{\vec{J}}_a+\hat{\vec{J}}_b$ is given by the largest possible $J_{\max}=M_{\max}=m_{a,\max}+m_{b,\max}=J_a+J_b$. There is exactly one largest (and smallest) M and thus exactly one sequence of states$|J_{\max},M\rangle$. In contrast, there are two ways of obtaining the second-largest $M_{\max}-1$. A linear combination of these two states $|m_{a,\max}-1,m_{b,\max}\rangle$ and $|m_{a,\max},m_{b,\max}-1\rangle$ gives the $M_{\max}-1$ state for $J_{\max}$, while the other, orthogonal, combination starts another sequence with $J=J_{\max}-1$. At least at first, one state in addition to the one 'needed' for the existing J becomes available each time M is lowered by one. One can see this by looking at the rectangle of all allowed (m_a,m_b) pairs. Each value of M then corresponds to a diagonal line with $m_b=M-m_a$. On thus finds total angular momenta $J_{\max}$, $J_{\max}-1$, ...

With such a rectangle, one can also see that diagonals stop growing at some point: when the corner of its shorter side (lower right or upper left) is reached. For $J_a\geq J_b$ ($J_b\geq J_a$), this happens for $M'=J_a-J_b$ ($M'=J_b-J_a$). For all M between $\pm M'$, we find $2J_b$ ($2J_a$) pairs of values $m_a+m_b=M$, and the number starts to shrink again for $M<-M'$. Consequently, no new J-sequences can be added. There are just enough states to complete at $-M'$ the sequence started at M', which belongs to the smallest total $J=|J_a-J_b|$.

The allowed values of J are thus J_a+J_b, J_a+J_b-1, J_a+J_b-2, ..., $|J_a-J_b|$ and for each value, there is exactly one ladder with all the allowed M.

The eigenstates can also be obtained in a similar manner as in the two examples. As they are expressed in terms of the original tensor-product basis $|m_a,m_b\rangle$, we need the coefficients. The states $|J_{\max},\pm M_{\max}\rangle$ can be chosen as $|\pm m_{a,\max},\pm m_{b,\max}\rangle$, i.e., real. Further coefficients then follow from the application of $J_\pm$ as above and consequently continue to be real.

(One could, if wanted, introduce complex phases, but it is clearly possible to keep coefficients real.) These coefficients are called Clebsch-Gordan coefficients, they can be obtained via recursion relations or from tables.

The basis transformation from the tensor-product basis to the $|J, M\rangle$ basis was motivated by the transition from conserved individual angular momenta to conserved total angular momentum. In terms of symmetry, we went from rotational symmetry for each subsystem (either individual particles or the spin and position-space sectors) to rotational symmetry of the combined system. It turns out that rotations of the total system, written as matrices, indeed look different in the two bases: In the tensor-product basis $|m_a, m_b\rangle$, rotations usually mix all states, an exception being rotation around the z axis. In the $|J, M\rangle$ basis, in contrast, the rotation matrix is block diagonal and only mixes states with the same J. It is moreover the "best possible" block diagonal form, i.e., it is not possible to find another basis transformation leading to smaller blocks. One then speaks of "irreducible" representations of the involved symmetry, here rotations, because the blocks cannot be made smaller. The total rotation matrix is called a "direct sum" of the submatrices corresponding to the individual J-values.

4 Mixed states

Le Bellac

So far, we have always used "pure" states, i.e., the quantum systems under consideration were in well defined quantum states. In this chapter, we will look at various situations that deal with a break down of this assumption.

4.1 Singlet state as an entangled state

In Sec. 3.4.1, two spins were added into a total angular momentum. In this process, states are written in a tensor-product basis (3.55). A tensor-product state can easily be "separated" again, because we can simply read off the state of each subsystem. In the states of (3.55), each spin is in a specific state defined by its magnetic quantum number. This changed, however, once we went to the singlet-triplet basis (3.68,3.69) that is more appropriate for interacting spins. The quantum numbers are here total spin and total magnetic quantum number M and two of the states (the ones with $M = 0$) are not tensor-product states. This also means that they cannot be similarly decomposed, when two spins are in a singlet state, we cannot say whether the first (or the second) of them is in state $|\uparrow\rangle$ or $|\downarrow\rangle$.

At first, we check that the singlet state $\frac{1}{\sqrt{2}}(|\uparrow\rangle \otimes |\downarrow\rangle - |\downarrow\rangle \otimes |\uparrow\rangle)$ is indeed no tensor-product state, not even one corresponding to a a different choice for the single-spin basis. To do this, we compare it to the most general two-spin tensor-product state

$$\begin{aligned}
&(\mu_1|\uparrow\rangle + \mu_2|\downarrow\rangle) \otimes (\lambda_1|\uparrow\rangle + \lambda_2|\downarrow\rangle) = \\
&\quad = \mu_1\lambda_1|\uparrow\rangle \otimes |\uparrow\rangle + \mu_1\lambda_2|\uparrow\rangle \otimes |\downarrow\rangle + \mu_2\lambda_1|\downarrow\rangle \otimes |\uparrow\rangle + \mu_2\lambda_2|\downarrow\rangle \otimes |\downarrow\rangle = \\
&\quad = c_{11}|\uparrow\rangle \otimes |\uparrow\rangle + c_{12}|\uparrow\rangle \otimes |\downarrow\rangle + c_{21}|\downarrow\rangle \otimes |\uparrow\rangle + c_{22}|\downarrow\rangle \otimes |\downarrow\rangle
\end{aligned} \tag{4.1}$$

Even without knowledge of the coefficients μ_1, μ_2, λ_1 and λ_2, the coefficients have to fulfill $c_{11} \cdot c_{22} = c_{12} \cdot c_{21}$ because both products are given by $= \mu_1\mu_2\lambda_1\lambda_2$. For the singlet state, in contrast, $c_{11} = c_{22} = 0$ and $c_{12} = -c_{21} = \frac{1}{\sqrt{2}}$ and the product cannot be equal. The singlet state thus has a form that cannot be achieved with any single tensor-product state. Such a combined state that cannot be separated into states of its subsystems is called *entangled*.

In the singlet state, the z component of the first spin, $S_{z,a}$ has expectation value 0, because the probabilities for $+\frac{\hbar}{2}$ and $-\frac{\hbar}{2}$ are equal, which would also be true if spin a were in a state $\frac{1}{\sqrt{2}}(|\uparrow\rangle + \mathrm{e}^{i\phi}|\downarrow\rangle)$. In the latter case, however, the spin has a finite expectation value along *some* direction, just one that is orthogonal to the z-axis. For $\phi = 0$, e.g., it is the state $|x,\uparrow\rangle$, i.e., the expectation value for S_x is maximal $\frac{\hbar}{2}$. Similarly, it would be found to be polarized along $+y$ for $\phi = \pi/2$. In both cases, the spin is thus as well defined as possible in quantum mechanics. In contrast, expectation values vanish along *any* direction when calculated for one of the spins in a singlet state.

4.2 Density matrices

As the probability of finding "up" is exactly 50% along any axis, the state of one spin in a singlet is not defined at all. In addition to this possibility and to the case of a polarized beam of spins, it is possible to have intermediate situations. E.g., we might have a beam where 70% of spins spin lie along $+z$ and 30% along $-z$. The two (or more) possibilites do not even have to be given by orthogonal states, we could also add 70% along $+z$ and 30% along $-y$.

We are here going to discuss how to treat such incomplete information. The formalism to treat cases where the state of the quantum system is not exactly known should naturally also be applicable to the previous case of pure states, which can be seen as a special case with one state having probability 100%. A plausible way to extend the calculation of expectation values to a mixture of states would be to first obtain them for each pure state of the mixture and then average the results using weights given by the probabilities. This is indeed very close to what is done, however, it is useful to introduce a slightly different formalism in order to have all the information about the quantum state encoded in one object.

When starting from pure states, we request that the all-important expectation values remain unchanged, but somewhat rewrite them

$$\langle A\rangle_\phi = \langle\phi|A|\phi\rangle = \sum_m \langle\phi|A|m\rangle\langle m|\phi\rangle = \sum_m \langle m|\phi\rangle\langle\phi|A|m\rangle = \sum_m \langle m|P_\phi A|m\rangle = \mathrm{tr}(P_\phi A)\ . \tag{4.2}$$

We have here gone from looking at the state $|\phi\rangle$ to considering the projector $P = |\phi\rangle\langle\phi|$ onto it. Projectors can also project onto a subspace (e.g. the one belonging to a degenerate eigenvalue) instead of a single state, they are Hermitian $P^\dagger = P$ and fulfill $P^2 = P$.

In the formalism for imperfectly polarized states, one now replaces the projector by a sum of projectors onto states $|\phi_i\rangle$ that are weighted by the probability $p(i)$ of finding the (sub)system in this state. This object

$$\hat\rho = \sum_i p(i)|\phi_i\rangle\langle\phi_i| \tag{4.3}$$

is usually called "density matrix", but "state operator/matrix" or "statistical operator" are also in use. The last designation seeks to emphasize that the averaging introduced here comes from a lack of information about the (sub)system and is in addition to the quantum mechanical uncertainty. Expectation values are then calculated as

$$\begin{aligned}\langle A\rangle_\rho = \mathrm{tr}\,(\rho A) = \sum_m \langle m|\rho A|m\rangle = \sum_m\sum_i \langle m|\phi_i\rangle p(i)\langle\phi_i|A|m\rangle = \\ = \sum_i p(i)\langle\phi_i|A\underbrace{\sum_m |m\rangle\langle m|}_{=\mathbb{I}}\phi_i\rangle = \sum_i p(i)\langle\phi_i|A|\phi_i\rangle\ ,\end{aligned} \tag{4.4}$$

and indeed are weighted means. All the information about the quantum system is contained in the matrix ρ (rather than the state $|\phi\rangle$ as in a pure system). The states $|\phi_i\rangle$ are not necessarily orthogonal to each other and the same matrix ρ can arise from different ways of combining states. The trace can be evaluated in any basis $|m\rangle$, which does not have to be the eigenbasis of ρ. We can infer a few properties of density matrices $\hat\rho$

- As a sum of Hermitian projectors (weighted with real numbers), ρ is Hermitian itself.
- Eigenvalues of ρ are not negative, and they correspond to probabilities $p(n)$ of finding the system in eigenstate $|n\rangle$ of ρ. As a consequence $\langle\psi|\rho|\psi\rangle \geq 0$ for arbitrary ψ.
- The trace, which is independent of the basis, is one.
- If and only if $\rho^2 = \rho$, then ρ describes a pure $|\phi\rangle$ state and ρ is a projector onto this state.

Even though ρ is an operator, it is time-dependent in the Schrödinger and not in the Heisenberg picture, because it inherits its time-dependence form the states. Using the Schrödinger equation, one finds

$$\frac{d}{dt}\rho = \sum_i p(i)\left(\frac{d|\phi_i\rangle}{dt}\langle\phi_i| + |\phi_i\rangle\frac{d\langle\phi_i|}{dt}\right) = \sum_i p(i)\left(-\frac{i}{\hbar}H|\phi_i\rangle\langle\phi_i| + \frac{i}{\hbar}|\phi_i\rangle\langle\phi_i|H\right) = -\frac{i}{\hbar}[H,\rho]\ . \tag{4.5}$$

4.2.1 Reduced density matrices

A mixture of states could arise because we do not know enough about the system, e.g., we might have a random or partially polarized beam of spins, because we were not able to polarize them fully. However, another source of mixtures was at play in the singlet state discussed above: Here, the full system was in a pure state $|J=0, M=0\rangle$ and only when we are interested in a subsystem, the entanglement lead to a mixture rather than a pure state *for this subsystem* of a perfectly known state.

The density matrix of the subsystem is then found by "tracing out" the states belonging to the rest of the system. We work in the tensor-product basis and states are labelled by pairs of indices (e.g. n_A, n_B or m_A, m_B) giving the basis states in each subsystem A or B. The expectation value of an operator $A = A \otimes \mathbb{I}$ acting on subsystem A should be obtained, and we find

$$\begin{aligned}
\langle A\otimes\mathbb{I}\rangle = \mathrm{tr}\left[(A\otimes\mathbb{I})\rho\right] &= \sum_{(n_A,n_B)}\langle n_A,n_B|(A\otimes\mathbb{I})\rho|n_A,n_B\rangle = \\
&= \sum_{(n_A,n_B)}\sum_{(m_A,m_B)}\langle n_A,n_B|(A\otimes\mathbb{I})|m_A,m_B\rangle\langle m_A,m_B|\rho|n_A,n_B\rangle = \\
&= \sum_{(n_A,n_B)}\sum_{(m_A,m_B)}\langle n_A|A|m_A\rangle\underbrace{\langle n_B|\mathbb{I}|m_B\rangle}_{=\delta_{n_B,m_B}}\langle m_A,m_B|\rho|n_A,n_B\rangle = \\
&= \sum_{n_A}\sum_{m_A}\langle n_A|A|m_A\rangle\sum_{n_B}\langle m_A,n_B|\rho|n_A,n_B\rangle\ ,
\end{aligned} \tag{4.6}$$

where we introduce the *reduced density matrix*

$$\hat{\rho}^A = \sum_{n_A,m_A}|n_A\rangle\rho^A_{n_A,m_A}\langle m_A| = \sum_{n_A,m_A}|n_A\rangle\left(\sum_{n_B}\langle n_A,n_B|\rho|m_A,n_B\rangle\right)\langle m_A| \tag{4.7}$$

resp.

$$\langle n_A|\hat{\rho}^A|m_A\rangle = \sum_{n_B}\langle n_A,n_B|\rho|m_A,n_B\rangle\ . \tag{4.8}$$

It then becomes

$$\langle A \otimes \mathbb{I} \rangle = \sum_{n_A} \sum_{m_A} \langle n_A|A|m_A\rangle\langle m_A|\hat{\rho}^A|n_A\rangle = \sum_{n_A} \langle n_A|A\hat{\rho}^A|n_A\rangle = \mathrm{tr}\left(A\rho^A\right) \tag{4.9}$$

The information about a subsystem is thus encoded in the operator (4.7). It is a density matrix:

- It is Hermitian, because ρ is, i.e. $\langle m_A, m_B|\rho|n_A, n_B\rangle = \langle n_A, n_B|\rho|m_A, m_B\rangle^*$, and consequently

$$\langle m_A|\hat{\rho}^A|n_A\rangle = \sum_{n_B} \langle m_A, n_B|\rho|n_A, n_B\rangle = \sum_{n_B} \langle n_A, n_B|\rho|m_A, n_B\rangle^* = \langle n_A|\hat{\rho}^A|m_A\rangle^* \tag{4.10}$$

- Its trace is that of ρ and thus one:

$$\sum_{n_A} \langle n_A|\hat{\rho}^A|n_A\rangle = \sum_{n_A, n_B} \langle n_A, n_B|\rho|n_A, n_B\rangle = 1 \tag{4.11}$$

- Elements on its diagonal $\langle n_A|\rho^A|m_A\rangle \geq 0$ if $\langle n_A, n_B|\rho|m_A, m_B\rangle \geq 0$, i.e., it is a positive matrix. As a further consequence, its eigenvalues are not negative.
- That ρ^A is a projector if and only if $(\rho^A)^2 = \rho^A$ follows from the previous points: In its eigenbasis, the eigenvalues of $(\rho^A)^2$ are those of ρ^A squared, i.e., $p(n_A)^2$. As $1 \geq p(n_A)^2 \geq 0$, $p(n_A)^2 \leq p(n_A)$ and equality holds only for 0 and 1. [1]

4.3 Entangled states, Bell's inequality and locality

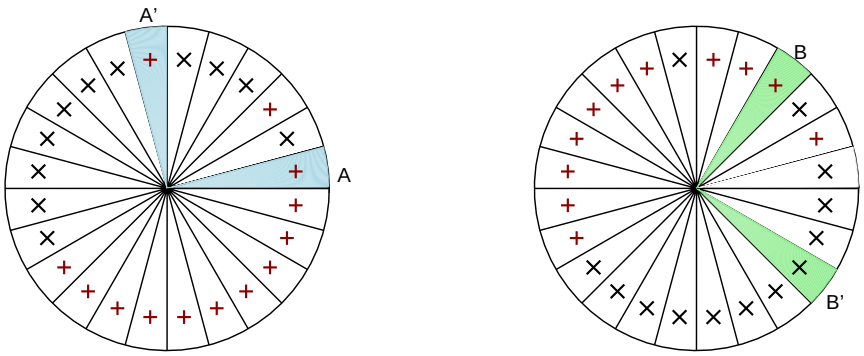
Closely following Le Bellac, the argument is based on a paper by Bell [2] on socks. Said paper is way more entertaining to read than the following discussion, but the experiment with the washing machines is less suitable for the class room than the paper version of Fig. 4.1.

If a particle with $J = 0$ and linear momentum $p = 0$ (i.e. at rest) decomposes into two spin$\frac{1}{2}$ particles, then these have to be in a total state of $J = 0$, i.e. singlet and to have opposite momentum $p_A = -p_B$. Assuming they fly far apart without interacting with their environment, the two partners in the singlet can then be very far apart with two experimentalists usually called Alice and Bob.

If Alice uses a Stern-Gerlach apparatus to measure the spin of her particle to be $m_a = +\frac{\hbar}{2}$, then Bob's must of course have $m_b = -\frac{1}{2}$. The same holds for a SG apparatus along any other direction, e.g., along x. The argument for "hidden variables" now states that Alice might only decide quite late which SG apparatus to use and that moreover her measurement surely cannot "determine" the outcome of another measurement that is so far away that light would not be able to get there in time. Consequently, Bob's spin must already have taken all the information with it that it needs to both determine its m_b along z and along x. Even though SG apparatuses may not be able to measure both S_x and S_z of one spin and even though quantum mechanics does not permit to calculate both, they should in principle be defined - but hidden.

[1] For operators in infinite dimensions rather than matrices in finite dimensions, the arguments are as usual more subtle, but still hold.

[2] J. S. Bell, *Bertlmann's socks and the nature of reality*, J. Phys. Colloq. **42**, C2 (1981).



(a) (b)

Figure 4.1: Advent-calendar like boxes for Alice and Bob, corresponding to the 'hidden variable' concept. Each drawer of each box contains a spin "up" or "down", and Alice's and Bob's results are always opposite for drawers corresponding to any given direction. The shading shows directions that can be used to find a contradiction between quantum mechanics and these advent calendars, even if rules are introduced that allow only one drawer to be opened in each experiment.

The situation could be like this: Alice and Bob each took a bag with a ball with them. They know that one of them has a red ball inside and the other a black one, but do not know who got which. As soon as Alice sees that hers is read, she knows that Bob has the black one. There is no quantum mechanical singlet state at work, the bags rather had the information all the time, even if Alice and Bob did not. At this level, the scenario with preexisting (but inaccessible) information and the quantum mechanical singlet are perfectly equivalent. The scenario with balls could moreover be extended to the analogue of SG apparatuses along several direction: Instead of one bag with one ball, Alice and Bob could each have been given a large box with many small drawers, like an advent calendar, that are labelled by direction: For each "direction", the boxes were prepared to always have "opposite" balls in Alice's and Bob's drawer, see Fig. 4.1.

"SG experiments" can be performed with this box of drawers, where $+1$ might be assigned to finding a red ball and -1 to a black one. The two of them could then perform many experiments where they both open one drawer each and note the result. Later, they meet and compare their notes. If they always openend the same drawer, their balls must by construction have had "opposite" color, and one finds a perfect anticorrelation and the product of their results is always -1. If they open different drawers, it is a priori not clear what their results would be and they should presumably depend on the relative orientation of the "direction" whose drawer they open. However, one can still repeat the experiment many times, take notes, calculate products of measurements taken at the same time, and do statistics on them. The hidden variables are here the balls in all the unopened drawers: Even if a rule may say that only one drawer can be opened in each experiment (in order to mimic the fact that SG experiments cannot give information about a state's spin along different axes), the information exists in principle and better experiments might be able to find it. Analogously, a more complete theory than quantum mechanics should also include

these hidden variables.

Each time, Alice can find $\epsilon_a = 1$ or $\epsilon_a = -1$, if she opens drawer a, and likewise $\epsilon_{a'} = \pm 1$ if she opens another drawer a'. For Bob, possible results are $\epsilon_b = \pm 1$ and $\epsilon_{b'} = \pm 1$. A few things are clear without knowing anything about the dependence of the correlations on the "direction" written onto the drawer, e.g. that product of their "measurements" is always ± 1 as well, only the expectation value collected after many repeats will reveal the actual correlation $\langle \epsilon_a \epsilon_b \rangle$. One could now get the idea to write down as an interesting experimental quantity

$$x = \epsilon_a \epsilon_b + \epsilon_a \epsilon_{b'} + \epsilon_{a'} \epsilon_b - \epsilon_{a'} \epsilon_{b'} \ . \tag{4.12}$$

Each time, each participant must only open one drawer, so Alice cannot measure ϵ_a and $\epsilon_{a'}$ at the same time. But as the drawer and the ball exist, one is certainly allowed to write down this quantity, even though it may not be measurable.

Alice and Bob could now do many repeated experiments. Alice chooses a or a' each time and Bob b or b'. If they do this often enough, they should get enough measurements of all four combinations in Eq. (4.12) to get reliable statistics. Taken separately, each lies between -1 and 1 and the total would then be between -4 and 4. However, even without knowing anything about the directions, we can restrict the possible average result even more. One an rewrite the quantity to

$$x = \epsilon_a \left(\epsilon_b + \epsilon_{b'} \right) + \epsilon_{a'} \left(\epsilon_b - \epsilon_{b'} \right) \ , \tag{4.13}$$

which can not be measured at once, because Alice and Bob only open one drawer each. However, we know without measuring that it must be between -2 and 2 for each *advent calendar* (or box of drawers), see Figs. 4.1(a) and 4.1(b). Its average over many boxes of drawers can consequently not be outside ± 2 either. This is a "Bell's inequality".

The argument of Einstein, Podolski and Rosen was that the situation of a spin should in fact be like the case of the boxes of drawers: As Alice's and Bob's SG tests along the same axis always are perfectly anticorrelated and as their regions in space are too far apart to permit Alices' experiment to determine Bob's, the outcome must somehow have been encoded from the start. Even if only one SG experiment at a time is allowed and even if Alice an Bob cannot predict its outcome, the results to all possible SG experiments are determined and hidden like the balls in the drawers. In this case, the average over many experiments (where only one of the four terms of (4.12) can be measured each time) must be the same as the average over many boxes of drawers, where Eq. (4.13) holds and restricts the possible results.

Now, we go the the predictions of quantum mechanics, which are for a single SG experiment the same as those of a red-ball-black-ball game. Quantum mechanically, we can calculate the expectation value for (4.12) as well, they are given by $\langle \epsilon_a \epsilon_b \rangle = -\cos\theta$, where θ is the angle between a and b (both lie in the x-z plane and have unit length). We then specifically select $a = e_z$, $a' = e_x$, $b = \frac{1}{\sqrt{2}}(e_x + e_z)$, and $b' = \frac{1}{\sqrt{2}}(-e_x + e_z)$. The total expectation value is then

$$\begin{aligned} \langle \epsilon_a \epsilon_b \rangle + \langle \epsilon_a \epsilon_{b'} \rangle + \langle \epsilon_{a'} \epsilon_b \rangle - \langle \epsilon_{a'} \epsilon_{b'} \rangle &= -\cos\frac{\pi}{4} - \cos\frac{-\pi}{4} - \cos\frac{-\pi}{4} + \cos\frac{-3\pi}{4} = \\ &= -\frac{3}{\sqrt{2}} + \frac{-1}{\sqrt{2}} = -2\sqrt{2} < -2 \end{aligned} \tag{4.14}$$

The quantum mechanical result has here be shown to violate the above inequality, and a singlet is in an important way different from the box of drawers. The only possibility for the difference is here that we assumed that *all* drawers (also ones that were not measured) contain predetermined ϵ and used this assumption to establish the limit ± 2 for (4.13). (We set the average over boxes of drawers to be the same as the average over measurements.) This gave us the wrong limit and can thus not be allowed: The quantum-mechanical result is incompatible with the notion that the results to unperformed experiments have any reality as "hidden variables".

5 Time-independent Approximation methods

Nolting (QM), Sakurai I

Only very few quantum-mechanical problems can be solved exactly and approximation methods are thus needed. Two common ones are variational and perturbative approaches.

5.1 Variational Principle

The variational principle makes use of the fact that the wave function ϕ, for which

$$\langle A\rangle_\phi = \frac{\langle\phi|A|\phi\rangle}{\langle\phi|\phi\rangle} \tag{5.1}$$

is extremal, is an eigenstate of A. Assuming that the spectrum of A is known and is bounded from below (i.e. its lowest eigenvalue is some finite A_0), this makes immediate sense, as an arbitrary state $|\phi\rangle$ can be expressed in terms of the eigenstates $|a_n\rangle$ as $|\phi\rangle = \sum_n c_n|a_n\rangle$ and

$$\langle A\rangle_\phi = \frac{\sum_{n,m} c_n^* c_m \langle a_n|A|a_m\rangle}{\sum_n |c_n|^2\langle a_n|a_n\rangle} = \frac{\sum_n |c_n|^2 A_n}{\sum_n |c_n|^2} \geq \frac{\sum_n |c_n|^2 A_0}{\sum_n |c_n|^2} = A_0\ . \tag{5.2}$$

To make use of this observation, one precedes as follows:

1. Select a space of "trial wave functions". This is usually guided by symmetries one wishes the solution to have or by known solutions to similar problems. In principle, there are two slightly different approaches:
   - Write the trial wave function as a superposition of fixed wave functions $|\phi\rangle = \sum_i f_i(\alpha_1, \alpha_2, \dots)|\psi_i\rangle$, where the parameters α entering the coefficients will be optimized.
   - Write a wave function that includes a variational parameter, e.g. the parameter α modifying the hydrogen wave function when it is used in the variational Ansatz $\mathrm{e}^{-\alpha r/a_b}$ for the helium atom.
2. Express the energy $\langle H\rangle_\phi = \langle\phi|H|\phi\rangle/\langle\phi|\phi\rangle$ as a function of the variational parameters α: $\langle H\rangle_\phi = h(\alpha_1, \alpha_2, \dots)$.
3. Minimize $h(\alpha_1, \alpha_2, \dots)$ with respect to the α_i. This would typically be done numerically.
4. Calculate the desired quantities using the approximate state obtained for the optimal α_i.

Of course, the quality of the solution will critically depend on the choice of trial functions, i.e., on how well this subspace overlaps with the true ground states. The problem is that this is hard to judge in practice, as one does after all not know the true ground state. An advantage of the variational principle is, however, that it strictly provides an *upper* bound

to the ground-state energy and can never go to values that are too small. If two different variational treatments give different energies, the lower result is thus automatically the better approximation for the ground-state energy.

If the ground state $|\phi_0\rangle$ (with energy H_0) of H is exactly known, a variational principle can also be used to find the lowest excited state by defining an operator

$$H_1 = H - |H_0|\phi_0\rangle\langle\phi_0| \tag{5.3}$$

and minimizing its expectation value. If $|\phi_0\rangle$ is only approximately known, this estimate is, however, no longer "strictly variational", which means that it can go below the first excited state.

5.2 Perturbation theory

The second widely used approximation method is perturbation theory, where one tries to decompose ones Hamiltonian into a part with known solution and a hopefully "small" rest:

$$H = H_0 + V, \quad \text{with known} \quad H_0|n^0\rangle = E_n^0|n^0\rangle \ . \tag{5.4}$$

The rest V is called the perturbation. While this is sometimes taken to be time-dependent, see Sec. 6.2 later, we start with time-independent perturbations.

5.2.1 Time-independent non-degenerate perturbation theory

The first case that we will investigate is the time-independent problem (as in this equation), where the ground state $|n^0\rangle$ of H_0 is assumed to be non-degenerate and to be separated from the rest of the spectrum by a finite gap. (Higher states can be degenerate, and there can also be a continuous part of the spectrum.)

In order to later be better able to keep track of the orders, in which the perturbation enters the results, one introduces an explicit factor λ into the Hamiltonian:

$$H = H(\lambda) = H_0 + \lambda V \tag{5.5}$$

We are now of course interested in the eigenstates and energies (particularly the lowest state) of the full Hamiltonian

$$(H_0 + \lambda V)|n\rangle = E_n|n\rangle = (E_n^0 + \Delta_n)|n\rangle \ , \tag{5.6}$$

where Δ_n is the change in the energy due to the perturbation. This can be rewritten to

$$(E_n^0 - H_0)|n\rangle = (\lambda V - \Delta_n)|n\rangle \ , \tag{5.7}$$

A first thing to note is that this equation "lives" in the Hilbert space orthogonal to the unperturbed ground state $|n^0\rangle$, because

$$\langle n^0|(E_n^0 - H_0)|n\rangle = 0, \quad \text{and thus also} \tag{5.8}$$

$$\langle n^0|(\lambda V - \Delta_n)|n\rangle = 0 \ . \tag{5.9}$$

The second equation will later give us a useful relation to determine Δ_n. In order to make the subspace orthogonal to $|n^0\rangle$ visible, we introduce the projector $Q_n = 1 - P_n = \mathbb{I} - |n^0\rangle\langle n^0|$. Multiplying (5.7) by Q_n gives the equation

$$Q_n(E_n^0 - H_0)|n\rangle = Q_n(\lambda V - \Delta_n)|n\rangle \ , \tag{5.10}$$

and within the subspace defined by Q_n, $(E_n^0 - H_0)^{-1}$ is well defined and finite, because E_n^0 was assumed to be non-degenerate. Moreover, when we apply $(E_n^0 - H_0)^{-1}$, we can commute it with Q_n and find

$$Q_n|n\rangle = (E_n^0 - H_0)^{-1} Q_n(\lambda V - \Delta_n)|n\rangle \ , \tag{5.11}$$

The unknown vector $|n\rangle = (P_n + Q_n)|n\rangle$ thus has a component parallel to $|n^0\rangle$, which is large for a small perturbation but for which (5.7) vanishes and is unhelpful, and an orthogonal part obeying (5.11), yielding

$$|n\rangle = P_n|n\rangle + Q_n|n\rangle = \underbrace{c(\lambda)}_{=1}|n^0\rangle + \frac{Q_n}{E_n^0 - H_0}(\lambda V - \Delta_n)|n\rangle \ . \tag{5.12}$$

The constant $c(\lambda)$ is chosen to be 1, i.e., $|n\rangle$ is not normalized to $\langle n|n\rangle = 1$ but to $\langle n^0|n\rangle = 1$. This equation still contains $|n\rangle$ on both sides, in addition to the unknown quantity $\Delta_n = E_n - E_n^0$, however, it can be used to obtain a systematic list of terms proportional to various orders of λ.

The aim is to obtain expansions

$$|n\rangle = |n^0\rangle + \lambda|n^1\rangle + \lambda^2|n^2\rangle + \dots \tag{5.13}$$

$$\Delta_n = \Delta^0 + \lambda\Delta_n^1 + \lambda^2\Delta_n^2 + \dots \tag{5.14}$$

The vectors $|n^i\rangle$ are found by inserting these expansions into (5.12) and collecting the powers of λ.

$$\begin{aligned} |n\rangle &= |n^0\rangle + \frac{Q_n}{E_n^0 - H_0}(\lambda V - \lambda\Delta_n^1 - \lambda^2\Delta_n^2 - \dots)\left(|n^0\rangle + \lambda|n^1\rangle + \lambda^2|n^2\rangle + \dots\right) = \\ &= |n^0\rangle + \lambda \underbrace{\frac{Q_n}{E_n^0 - H_0}(V - \Delta_n^1)|n^0\rangle}_{=|n^1\rangle} + \lambda^2 \underbrace{\frac{Q_n}{E_n^0 - H_0}\left((V - \Delta_n^1)|n^1\rangle - \Delta_n^2|n^0\rangle\right)}_{=|n^2\rangle} + \dots \end{aligned} \tag{5.15}$$

Since the projector Q_n removes any component parallel to $|n^0\rangle$, the vectors $|n^i\rangle$ are orthogonal to $|n^0\rangle$ and the part $\lambda^k\Delta_n^k|n^0\rangle$ in fact drops out.

For the energy corrections Δ_n, the equation to look at is (5.9), which gives

$$\begin{aligned} \Delta_n \underbrace{\langle n^0|n\rangle}_{=1} = \lambda\langle n^0|V|n\rangle &= \lambda\langle n^0|V\left(|n^0\rangle + \lambda|n^1\rangle + \lambda^2|n^2\rangle + \dots\right) = \\ &= \lambda \underbrace{\langle n^0|V|n^0\rangle}_{=\Delta_n^1} + \lambda^2 \underbrace{\langle n^0|V|n^1\rangle}_{=\Delta_n^2} + \dots \end{aligned} \tag{5.16}$$

There is no zeroth-order contribution to Δ_n, because E_n^0 is already the full energy in zeroth-order.[1]

The first few correction are given by

$$\Delta_n^1 = \langle n^0|V|n^0\rangle \tag{5.17}$$

$$|n^1\rangle = \frac{Q_n}{E_n^0 - H_0}(V - \langle n^0|V|n^0\rangle)|n^0\rangle = \frac{Q_n}{E_n^0 - H_0}V|n^0\rangle = \sum_{k\neq n}|k^0\rangle\frac{\langle k^0|V|n^0\rangle}{E_n^0 - E_k^0} \tag{5.18}$$

$$\Delta_n^2 = \langle n^0|V\frac{Q_n}{E_n^0 - H_0}V|n^0\rangle = \sum_{k\neq n}\frac{|\langle n^0|V|k^0\rangle|^2}{E_n^0 - E_k^0} = -\sum_{k\neq n}\frac{|\langle n^0|V|k^0\rangle|^2}{E_k^0 - E_n^0} \leq 0 \tag{5.19}$$

$$\begin{aligned}|n^2\rangle &= \frac{Q_n}{E_n^0 - H_0}(V - \Delta_n^1)|n^1\rangle = \\ &= \sum_{l\neq n}\sum_{k\neq n}|l^0\rangle\frac{\langle l^0|V|k^0\rangle}{E_n^0 - E_l^0}\frac{\langle k^0|V|n^0\rangle}{E_n^0 - E_k^0} - \langle n^0|V|n^0\rangle\sum_{k\neq n}|k^0\rangle\frac{\langle k^0|V|n^0\rangle}{(E_n^0 - E_k^0)^2}\end{aligned} \tag{5.20}$$

Generally, the perturbed states are mixtures of the unperturbed eigenstates, which does not impose fundamental restriction, as they form a basis. If the series converges, one can hope that perturbation theory is valid. A criterion for "small" can be guessed from observing that each order adds factors like $\frac{\langle n^0|V|k^0\rangle}{E_n^0 - E_k^0}$: Requiring this factor to be small amounts to requiring that a perturbation be small compared to the energy gap separating the state of interest (usually the ground state) from other unperturbed states.

5.2.2 Time-independent degenerate perturbation theory

If the ground state, or the state of interest, is degenerate, the situation changes somewhat, because the denominator $1/(E_n^0 - H_0)$ is then not finite for all states $\neq |n^0\rangle$. One proceeds in a similar manner as before, but the projector Q_n then projects only onto the subspace with different energy and one introduces its complement, projector P_n onto the degenerate subspace. To reduce the number of sub- and superscripts, they are now called P and Q and if indices $\leq g$ designates the degenerate states, they are

$$P = \sum_{m\leq g}|m_g^0\rangle\langle m_g^0| \quad \text{and} \quad Q = \sum_{k>g}|k_g^0\rangle\langle k_g^0| = \mathbb{I} - P\,. \tag{5.21}$$

One again starts from Eq (5.7), but instead of first investigating what happens upon multiplication with $\langle n^o|$, one now writes two equations obtained by multiplying with P and Q:

$$P(E_n^0 - H_0)|n\rangle = 0 = P(\lambda V - \Delta_n)|n\rangle \tag{5.22}$$

and

$$Q(E_n^0 - H_0)|n\rangle = Q(\lambda V - \Delta_n)|n\rangle \quad \Rightarrow \quad Q|n\rangle = \frac{Q}{E_n^0 - H_0}(\lambda V - \Delta_n)|n\rangle\,. \tag{5.23}$$

[1] If one wants, one can go back to E_n and take E_n^0 as its zero-order part, its higher-order parts are then given by Δ_n.

The first of these, (5.22), gives again the equation for the energy correction Δ_n, as its counterpart (5.9),

$$\Delta_n P|n\rangle = \lambda PV|n\rangle = \lambda PVP|n\rangle + \lambda PVQ|n\rangle \ . \tag{5.24}$$

This is considerably more annoying than (5.9), because it is an equation for a vector in the subspace $m \leq g$ and not for a single number. We are going to see that we accordingly end up with matrix equations, where eigenvalues determine the energy corrections and eigenvectors give (parts of) the perturbation-theory eigenstates.

5.2.2.1 First-order degenerate perturbation theory

Any contributions $Q|n\rangle$ to the approximate vector $|n\rangle$ carry at least one factor λ, i.e., cannot be part of the *lowest-order* approximation to Δ_n. The first-order in λ for Δ_n is consequently given by

$$\Delta_n^1 P|n\rangle = PVP|n\rangle \ , \tag{5.25}$$

which turns out to be an eigenvalue equation restricted to the sector $m \leq g$. As the dimension g of the degenerate subspace is much smaller than that of the full Hilbert space, one can hope to be able to solve this equation and to diagonalize V in this subspace.

Since the unperturbed Hamiltonian H_0 is degenerate for $m \leq g$, its eigenstates in this subspace are not uniquely defined and we can choose any basis within the subspace. A smart choice for a basis is the one where PVP is diagonal. (But this choice is by no means necessary, the case of "almost degenerate" levels can be discussed without this freedom.) The advantage of using the eigenbasis of PVP is convenience, because the first-order energy correction becomes diagonal:

$$\Delta_m^1 \langle m^0|P|n\rangle = \Delta_m^1 = \langle m^0|V|m^0\rangle \ . \tag{5.26}$$

It should be noted that the correction can depend on m, we would in fact expect this to happen if the degeneracy of H_0 is not due to a symmetry common to H_0 and V. In this case, the degeneracy is (at least partially) lifted by the perturbation V and (some of the) vectors $|m^0\rangle$ are uniquely defined.

Especially if the lowest-energy state is now unique, one can think of three possibilities:

1. Stop the perturbation theory here, in first order. The result is that perturbation V lifts the degeneracy of H_0 (and possibly selects a ground state). This can be a valid and quite interesting result.
2. Note that the ground state is now no longer degenerate and continue with non-degenerate perturbation theory. However, this approach more popular in textbooks (Nolting (QM), Sakurai I) than practically relevant: The energy separating the newly found ground state from other, formerly degenerate, states is on the order of V, which cannot be assumed to be large compared to the perturbation V.
3. Continue with a "degenerate" treatment that treats all states of the originally degenerate subspace in an equivalent way. Such an approach can in fact even be used if the lowest eigenstates of H_0 are not degenerate, but very close in energy and separated from higher-lying states by a gap large compared to V.

We will here continue with the last option, degenerate perturbation theory for states that are not necessarily degenerate, but may only be close in energy.

5.2.2.2 Higher-order degenerate perturbation theory

Concerning the corrections to the eigenvectors, it is helpful to discuss the subspaces P (i.e., the originally degenerate ground-state space) and Q (the orthogonal space) separately. Going from an arbitrary initial basis for P to the eigenbasis of V, see (5.25), can here be seen as a 'first-order correction' within P. For the orthogonal space Q, we find the first-order correction from (5.23), where we again use that $Q|n\rangle$ would necessarily contribute another factor of λ:

$$\begin{aligned} Q|n^1\rangle &= \frac{Q}{E_n^0 - H_0}(V - \Delta_m^1)|n^0\rangle = \frac{Q}{E_n^0 - H_0}(V - \Delta_m^1)P|n^0\rangle = \\ &= \frac{Q}{E_n^0 - H_0}VP|n^0\rangle - \frac{Q}{E_n^0 - H_0}PVP|n^0\rangle = \\ &= \sum_{k>g}\sum_{m\leq g}|k^0\rangle\frac{\langle k^0|V|m^0\rangle}{E_n^0 - E_k}\langle m^0|n^0\rangle = \sum_{k>g}|k^0\rangle\frac{\langle k^0|V|n^0\rangle}{E_n^0 - E_k}\ , \end{aligned} \tag{5.27}$$

Where $PVP|n\rangle$ was inserted for $\Delta_m^1 P|n\rangle$, see (5.25) and where $QP = 0$ makes the second term drop out. This looks just like in the non-degenerate case, except that the statement is restricted to $Q|n\rangle$ and that more states are excluded from the sum over k. [2] A further difference is that we might evaluate this correction to several (or all) $|m^0\rangle$ if we are interested in the whole low-energy subspace or if the ground state is still degenerate. The second-order energy correction might now be evaluated from (5.24) as $\Delta_n^2 P|n\rangle = PV|m^1\rangle = PVP|m^1\rangle + PVQ|m^1\rangle$. However, it turns out to be cleaner and simpler not to focus on a single low-energy state $|m^1\rangle$, but to consider the whole low-energy subspace at once.

Going back to (5.24), we remember that both Δ and $Q|n\rangle$ carry at least one factor of λ and thus neglect corrections *higher than* second order, i.e., we discuss first and second order together:

$$\begin{aligned} \Delta_n P|n\rangle &= \lambda PV|n\rangle = \lambda PVP|n\rangle + \lambda PVQ|n\rangle = \lambda PVP|n\rangle + \lambda PV\frac{Q}{E_n^0 - H_0}(\lambda V - \Delta_n)|n\rangle = \\ &= \lambda PVP|n\rangle + \lambda^2 PV\frac{Q}{E_n^0 - H_0}V(P|n\rangle + \underbrace{Q|n\rangle}_{\propto\lambda}) - \lambda PV\frac{1}{E_n^0 - H_0}\underbrace{\Delta_n}_{\propto\lambda}\underbrace{Q|n\rangle}_{\propto\lambda} \\ \Rightarrow\quad \Delta_n^{1+2}P|n\rangle &= P\left(V + V\frac{Q}{E_n^0 - H_0}V\right)P|n\rangle \end{aligned} \tag{5.28}$$

This is again an eigenvalue equation for a vector in the originally degenerate low-energy $g \times g$ subspace, with the eigenvalues Δ_n^{1+2} yielding the combined first- and second-order energy correction. No assumptions have here to be made about corrections included (or not) for $P|n\rangle$, as the eigenvalue equation "lives" in the full subspace given by P. The $g \times g$ matrix to be diagonalized is the sum of the first-order matrix

$$\Delta_{\tilde{m},m}^1 = \langle\tilde{m}|V|m\rangle\ , \tag{5.29}$$

as in (5.26), and a second-order contribution with matrix elements

$$\Delta_{\tilde{m},m}^2 = \sum_{k>g}\frac{\langle\tilde{m}^0|V|k^0\rangle\langle k^0|V|m^0\rangle}{E_n^0 - E_k}\ . \tag{5.30}$$

[2] Formally, we could extend the sum to include the other states with $\tilde{m} \leq g$, if we work in the eigenbasis of V, because the matrix elements then $\langle\tilde{m}^0|V|m^0\rangle$ vanish by construction.

Δ^1 and Δ^2 wil in general not commute, they have to be diagonalized together. The perturbed eigenvectors have a component in $Q|n\rangle$, whose first-order part is given by (5.27), and a component in $P|n\rangle$ given by the eigenvectors of the Δ-matrix. Normalization is here such that $||P|n\rangle|| = 1$, analogous to $\langle n^0|n\rangle = 1$ of the non-degenerate case.

5.2.2.3 Nearly-degenerate perturbation theory

Sometimes, perturbation theory is performed for "almost degenerate" states. Here, the basis of the $g \times g$ subspace can not simply be chosen to diagonalize first-order correction Δ^1, because the basis is already fixed by the not-quite-degenerate eigenstates of H_0. This case is treated as the one here, the combined matrix equation in the low-energy subspace is

$$\Delta_n^{0+1+2} P|n\rangle = P\left[\left(\sum_m (E_m^0 - E_0)|m^0\rangle\langle m^0|\right) + V + V\frac{Q}{E_0 - H_0}V\right]P|n\rangle \ , \tag{5.31}$$

where $\Delta^0 = \sum_m (E_m^0 - E_0)|m^0\rangle\langle m^0|$ is the "zeroth-order" correction given by the small but finite original energy differences. Higher-order correction to the energies are in general matrices in the $g \times g$ low-energy subspace as well. These matrices together are usually called an "effective low-energy Hamiltonian".

Corrections to the eigenstates arise in two ways: in $Q|n\rangle$ through the "obvious" way of (5.27) and in $P|n\rangle$ through the requirement that $P|n\rangle$ diagonalize the effective Hamiltonian. In practice, one is usually more interested in the part $P|n\rangle$.

5.2.2.4 Schrieffer-Wolff transformation

As we have seen above, degenerate or almost-degenerate perturbation theory leads to non-trivial approximate eigenstates even within the ground-state, resp. low-energy, manifold. In fact, the solution to an eigenvalue equation like (5.31), i.e. eigenstates $P|n\rangle$ of a low-energy effective Hamiltonian, are often more interesting than components $Q|n\rangle$ outside the low-energy space.

An example would be the Pauli equation, which describes electrons with spin and arises in a perturbation theory for the Dirac equation, see Sec. 9.3.2.2. While the Pauli equation starts from the relativistic Dirac equation, it lives as a second-order perturbation in the low-energy Hilbert space of non-relativistic electrons. (The large energy scale is here the rest mass.) Other examples are the Kondo model as a perturbative treatment of the Anderson-impurity Hamiltonian, the t-J model for the Hubbard model, the Kugel-Khomskii models for multi-orbital models, effective electron-electron attraction mediated by phonons in superconductors ...

In all these cases, interest focuses on the subspace reached by P, i.e., one tries to find vectors $P|n\rangle$ that optimally capture the effect of the perturbation onto the low-energy Hilbert space. We are effectively looking for a basis transformation acting within the P subspace. This philosophy is followed by the Schrieffer-Wolff approach to degenerate perturbation theory, also knwon as the FoldyWouthuysen transformation inthe context of the Dirac equation. One here tries to find a unitary transformation $e^{i\tilde{S}}$, with Hermitian $\tilde{S}$:

$$\tilde{H} = e^{i\tilde{S}} H e^{-i\tilde{S}} = H + [i\tilde{S}, H] + \frac{1}{2}\left[i\tilde{S}, [i\tilde{S}, H]\right] + \cdots = H_0 + \lambda V + i[\tilde{S}, H_0] + i\lambda[\tilde{S}, V] + \mathcal{O}(\lambda^2) \tag{5.32}$$

The requirement that $\tilde{H} \to H_0$ for $\lambda \to 0$ implies that $[\tilde{S}, H_0] \to 0$ for $\lambda \to 0$ and suggests that $\tilde{S} = \lambda S$ has to carry a factor of (at least) λ as well. The criterion for the transformation S is that it should remove as much of the coupling between low- and high-energy Hilbert spaces as possible. Such a coupling is mediated by part of the perturbation $V' = V - PVP - QVQ = QVP + PVQ$ in the original Hamiltonian, and the dominant contribution is the lowest, i.e., the first order. Incorporating the remaining 'harmless' part of the perturbation into $H_0' = H_0 + \lambda(PVP + QVQ)$, vanishing of first-order coupling

$$\tilde{H} = e^{i\lambda S} H \mathrm{e}^{-i\lambda S} = H_0' + \lambda \underbrace{\left(V' + i[S, H_0']\right)}_{=0} + \lambda^2 \left(i[S, V'] - \frac{1}{2}\left[S, [S, H_0']\right] \right) + \mathcal{O}(\lambda^3) =$$

$$= H_0' + \lambda^2 \frac{i}{2}\left[S, V'\right] + \ldots \tag{5.33}$$

yields $V' + i[S, H_0'] = 0$. The resulting operator S is then used to express the higher-order corrections via commutators.

$$iV' = [S, H_0'] = i \sum_{k,m} |k^0\rangle\langle k^0|V'|m^0\rangle\langle m^0| = \sum_{k,m} (E_m'^0 - E_k'^0)|k^0\rangle\langle k^0|S|m^0\rangle\langle m^0|$$

$$\Rightarrow \quad S = i \sum_{\substack{k>g \\ m\leq g}} |k^0\rangle \frac{\langle k^0|V'|m^0\rangle}{E_m'^0 - E_k'^0}\langle m^0| - i \sum_{\substack{k>g \\ m\leq g}} |m^0\rangle \frac{\langle m^0|V'|k^0\rangle}{E_m'^0 - E_k'^0}\langle k^0| \tag{5.34}$$

$$\tag{5.35}$$

Keeping only the part relevant to the low-energy subspace, the second-order term is given by

$$\frac{i}{2} P\left[S, V'\right] P = \frac{1}{2} \sum_{\substack{k>g \\ m,n\leq g}} |m^0\rangle \frac{\langle m^0|V'|k^0\rangle}{E_m'^0 - E_k'^0}\langle k^0|V'|n^0\rangle\langle n^0| + \frac{1}{2} \sum_{\substack{k>g \\ m,n\leq g}} |m^0\rangle\langle m^0|V'|k^0\rangle \frac{\langle k^0|V'|n^0\rangle}{E_n'^0 - E_k'^0}\langle n^0| \,.$$

For $|E_n'^0 - E_m'^0| = |\langle n^0|V|n^0\rangle - \langle m^0|V|m^0\rangle| \ll E_k'^0 - E_n'^0 \approx\ll E_k^0 - E_n^0$, this expression indeed becomes equivalent to (5.30). The difference between the two formulas carries another factor of $\langle V\rangle/|E_k^0 - E_n^0|$, i.e., it is of *third* order, so that both expressions can be "correct up to second order". The Schrieffer-Wolff variant is occasionally easier to evaluate, naemly in cases where we have much information on commutator relations.

6 Time dependence: Heisenberg picture and perturbation theory

6.1 Heisenberg picture

So far, time dependence has been assumed to affect the states, but this is not the only way to see things: only the time dependence of measurable quantities, like expectation values, has to be uniquely defined. In such an expectation value, time dependence can be shifted onto the operator as well:

$$\begin{aligned}\langle A\rangle_t &= \langle\psi(t)|A|\psi(t)\rangle = \langle U(t,t_0)\psi(t_0)|A|U(t,t_0)\psi(t_0)\rangle = \\ &= \langle\psi(t_0)|\underbrace{U^\dagger(t,t_0)AU(t,t_0)}_{A_H(t)}|\psi(t_0)\rangle = \langle A_H(t)\rangle_{t_0}\end{aligned} \tag{6.1}$$

While the first expression calculates the expectation value using the time-dependent wave function at time t, the wave function at t_0 is used in the second line, while the time-dependence has been shifted onto the operator. This is the Heisenberg picture and operators ar thus often designated with H. The time dependence on the wave function is the Schrödinger picture, operators are sometimes designated by an S.

An important simplification in using Heisenberg-type operators is that their product, e.g. in a commutator can be calculated in the Schrödinger picture and only the final operator given a time dependence:

$$\begin{aligned}[A_H(t),B_H(t)] = &\quad U^\dagger(t,t_0)A_SU(t,t_0)U^\dagger(t,t_0)B_SU(t,t_0) \\ &- U^\dagger(t,t_0)B_S\underbrace{U(t,t_0)U^\dagger(t,t_0)}_{=1}A_SU(t,t_0) = \\ &= U^\dagger(t,t_0)(A_SB_S - B_SA_S)U(t,t_0) = U^\dagger(t,t_0)[A_S,B_S]U(t,t_0)\end{aligned} \tag{6.2}$$

The states $|\dots\rangle$ are now time independent, but let us look at the time dependence of operators:

$$\begin{aligned}\frac{d}{dt}A_H(t) &= \frac{d}{dt}U^\dagger(t,t_0)A_SU(t,t_0) = \\ &= \left(\frac{d}{dt}U^\dagger(t,t_0)\right)A_SU(t,t_0) + \underbrace{U^\dagger(t,t_0)\frac{\partial A_S}{\partial t}U(t,t_0)}_{\frac{\partial A_H(t)}{\partial t}} + U^\dagger(t,t_0)A_S\left(\frac{d}{dt}U(t,t_0)\right)\end{aligned} \tag{6.3}$$

The time evolution of the unitary operators $U(t,t_0)$ is given by Eq. (1.30). As discussed above, $\partial_t U$ and dU/dt are equivalent as long as it is clear that the second time t_0 does not carry any t-dependence and does not enter the differentiation. In the case of the

(Schrödinger) operator A_S, $\partial_t A_S$ denotes an "explicit" time dependence, e.g., if something is switched on of off. Since operators in the Schrödinger picture do not carry any "implicit" time evolution, $\partial_t A_S = dA_S/dt$. dA_H/dt, on the other hand, includes the time evolution due to quantum dynamics in addition to the explicit change. The two symbols $\partial A_H/\partial t$ and dA_H/dt thus really have different meaning, the first is used to denote the Heisenberg (=time dependent) variant of $\partial_t A_S$.

Taking $\frac{d}{dt}U(t,t_0) = -\frac{i}{\hbar}H_S U(t,t_0)$ from Eq. (1.30), we find

$$\begin{aligned}\frac{d}{dt}A_H(t) &= \frac{i}{\hbar}\left(U^\dagger(t,t_0)H_S A_S U(t,t_0) - U^\dagger(t,t_0)A_S H_S U(t,t_0)\right) + \frac{\partial A_H(t)}{\partial t} \\ &= \frac{i}{\hbar}U^\dagger(t,t_0)[H_S,A_S]U(t,t_0) + U^\dagger(t,t_0)\frac{dA_S}{dt}U(t,t_0) = \\ &= \frac{i}{\hbar}[H_H,A_H] + \frac{\partial A_H(t)}{\partial t} \end{aligned} \tag{6.4}$$

If the Hamiltonian has no "explicit" time dependence, i.e., no time dependence in the Schrödinger picture, then it commutes with U and $U^\dagger$ and $H_H = H_S$. The "explicit" time dependence $\frac{\partial A_H(t)}{\partial t}$ is a time dependence in the Schrödinger picture, however, the operator describing dA_S/dt has to be translated into Heisenberg form.

6.1.1 Example: Harmonic oscillator

As an illustration, let us look at the harmonic oscillator, in particular at the time dependence of the anihilation operator. In the Schrdinger picture, a does not have any time dependence, the time evolution for operators given in (6.4) leads thus to

$$\frac{d}{dt}a_H(t) = \frac{i}{\hbar}[H_H, a_H(t)] = \frac{i}{\hbar}[H,a]_H \ , \tag{6.5}$$

where $[H,a]_H$ signals that we may calculate the commutator in the Schrödinger representation and transform the result into the Heisenberg picture.

$$\begin{aligned}[H,a]_H = U^\dagger[H_S,a_S]U = U^\dagger[\hbar\omega(\hat{N}+\frac{1}{2}),a]U = U^\dagger[\hbar\omega\hat{N},a]U = \hbar\omega U^\dagger[a^\dagger a,a]U = \\ = \hbar\omega U^\dagger(a^\dagger \underbrace{[a,a]}_{0} + \underbrace{[a^\dagger,a]}_{-1}a)U = -\hbar\omega U^\dagger a U = -\hbar\omega a_H(t)\end{aligned} \tag{6.6}$$

and consequently

$$\frac{d}{dt}a_H(t) = -i\omega a_H(t) \ . \tag{6.7}$$

This can be integrated to

$$a_H(t) = \mathrm{e}^{-i\omega t}a_H(t=0) = \mathrm{e}^{-i\omega t}a \ , \tag{6.8}$$

i.e., $a_H(t)$ has a time dependence analogous to that of classical variable $z(t)$ discussed in Sec. 1.5.3.

6.2 Time-dependent perturbation theory

Sakurai I

Time-dependent perturbation theory addresses somewhat different issues from the time-independent variant presented in Sec. 5.2: Now, it is assumed that we know the eigenstates of the (full) system of interest and the "small perturbation" is an additional time-dependent term,

$$H(t) = H_0 + V(t) \; . \tag{6.9}$$

A typical situation to be treated in this way is the impact of an experimental probe (depending on time) onto a known system. Exactly solvable time-dependent problems are pretty much restricted to the two-level problem, see exercises. Perturbation theory can help, if the perturbation is "small" – in the sense of "does not mess up the states of the system too much" – either due to small matrix elements $\langle n|V(t)|m\rangle$ or by virtue of being short-lived.

A typical problem treated with time-dependent perturbation theory is the reaction of a system to a perturbation that is "switched on" at time t_0. For $t \leq t_0$, it is usually found in an eigenstate $|i\rangle$ of H_0, but $|i\rangle$ is no longer an eigenstate for $t > t_0$ and consequently no longer stationary. One is then interested in the probability of finding the system in some other state, e.g. in other eigenstates $|n\rangle$ of H_0.

6.2.1 "Interaction" = Dirac picture

We are typically interested in the time-evolution of some initial state, usually an eigenstate of the unperturbed contant term H_0, once the perturbation is switched on. The answer is given by the Schrödinger time-evolution operator (1.25) applied to the initial state, however, we cannot integrate (1.30) to obtain it, because $V(t)$ and H_0 will not generally commute, so that the full Hamiltonian at time t does not commute with that at $t' \neq t$. It turns out to be more practical not to use the Schrödinger or Heisenberg pictures introduced so far, but to go to yet another representation, the Dirac or interaction picture. "Interaction" refers here to the prime application of describing a system interacting with some external and time-dependent field. We will not be able to integrate the new time-evolution operator either, but we will be able to obtain a perturbative expansion in terms of "small" $V(t)$, so that we can at least systematically approximate it.

In this new picture, the time dependence of states is given by

$$|\phi(t,t_0)\rangle_D = \mathrm{e}^{\frac{i}{\hbar}H_0(t-t_0)}|\phi(t,t_0)\rangle_S = \mathrm{e}^{\frac{i}{\hbar}H_0(t-t_0)}\hat{U}_S(t,t_0)|\phi(t_0,t_0)\rangle_S \; , \tag{6.10}$$

where $|\phi(t,t_0)\rangle_S$ and $|\phi(t,t_0)\rangle_D$ denote states in the Schrödinger and Dirac pictures at time t. As expectation values must not change when going from one picture to the other, i.e., ${}_S\langle\phi(t,t_0)|A_S|\phi(t,t_0)\rangle_S = {}_D\langle\phi(t,t_0)|A_D|\phi(t,t_0)\rangle_D$, operators in the Dirac picture have to evolve as

$$A_D(t,t_0) = \mathrm{e}^{\frac{i}{\hbar}H_0(t-t_0)}A_S\mathrm{e}^{-\frac{i}{\hbar}H_0(t-t_0)} \; , \tag{6.11}$$

for operators $A = A_S$ not depending on time. (Operators without subscripts are supposed to be in the Schrödinger picture.) This equation looks rather similar to the Heisenberg

picture, see (6.4), but the crucial difference is that the time dependence of operators is only given by H_0, not by the full hamiltonian H.

The time evolution of Dirac states can be obtained from that of Schrödinger states:

$$\begin{aligned} i\hbar\frac{d}{dt}|\phi(t,t_0)\rangle_D &= i\hbar\frac{d}{dt}\left(\mathrm{e}^{\frac{i}{\hbar}H_0(t-t_0)}|\phi(t,t_0)\rangle_S\right) = \\ &= i\hbar\frac{i}{\hbar}H_0\mathrm{e}^{\frac{i}{\hbar}H_0(t-t_0)}|\phi(t,t_0)\rangle_S + \mathrm{e}^{\frac{i}{\hbar}H_0(t-t_0)}\underbrace{i\hbar\frac{d}{dt}|\phi(t,t_0)\rangle_S}_{=H(t)|\phi(t,t_0)\rangle_S,\ \text{see(1.20)}} = \\ &= \mathrm{e}^{\frac{i}{\hbar}H_0(t-t_0)}\left(-H_0+H(t)\right)|\phi(t,t_0)\rangle_S = \mathrm{e}^{\frac{i}{\hbar}H_0(t-t_0)}V(t)|\phi(t,t_0)\rangle_S = \\ &= \underbrace{\mathrm{e}^{\frac{i}{\hbar}H_0(t-t_0)}V(t)\mathrm{e}^{-\frac{i}{\hbar}H_0(t-t_0)}}_{=V_D(t)}\underbrace{\mathrm{e}^{\frac{i}{\hbar}H_0(t-t_0)}|\phi(t,t_0)\rangle_S}_{=|\phi(t,t_0)\rangle_D} = V_D(t)|\phi(t,t_0)\rangle_D \end{aligned} \tag{6.12}$$

The dynamics of states are thus given by the perturbation $V(t)$ in this picture, while that of operators is given by H_0.

In a manner completely analogous to Sec. 1.2.2, one can now introduce a unitary operator $U_D(t,t_0)$ describing the time evolution of

$$|\phi(t,t_0)\rangle_D = U_D(t,t_0)|\phi(t_0,t_0)\rangle_D\ . \tag{6.13}$$

By inserting the definition of $|\phi\rangle_D$, one could relate U_D to the time-evolution operator U of the Schrödinger picture, which will here depend on both t and t_0, due to the time-dependence of $H(t)$. Rather than doing this, we proceed here again as in Sec. 1.2.2 and derive the time-derivative of U_D as

$$i\hbar\partial_t U_D(t,t_0) = V_D(t)U_D(t,t_0)\ . \tag{6.14}$$

The initial condition completing the differential equation is $U_D(t_0,t_0)=\mathbb{I}$.

6.2.2 Dyson series: Perturbation series for the time evolution

Due to the time dependence of $V_D(t)$, which is both due to (6.11) and contained in $V_S(t)$, one cannot integrate differential equation (6.14)to something like $\mathrm{e}^{-i/\hbar V_D(t-t_0)}$. One can, however, formally integrate it to obtain an implicit equation [1]

$$U_D(t,t_0) = 1 + \frac{1}{i\hbar}\int_{t'=t_0}^{t} dt' V_D(t')U_D(t',t_0)\ . \tag{6.15}$$

The use of the integral equation becomes clearer after repeating the approach:

$$\begin{aligned} U_D(t,t_0) &= 1 + \frac{1}{i\hbar}\int_{t'=t_0}^{t} dt' V_D(t')U_D(t',t_0) = \\ &= 1 + \frac{1}{i\hbar}\int_{t'=t_0}^{t} dt' V_D(t')\left(1 + \frac{1}{i\hbar}\int_{t''=t_0}^{t'} dt'' V_D(t'')U_D(t'',t_0)\right) = \\ &= 1 + \frac{1}{i\hbar}\int_{t'=t_0}^{t} dt' V_D(t') + \left(\frac{1}{i\hbar}\right)^2\int_{t'=t_0}^{t} dt'\int_{t''=t_0}^{t'} dt'' V_D(t')V_D(t'')U_D(t'',t_0)\ , \end{aligned} \tag{6.16}$$

[1] Taking the derivative and setting $t=t_0$ indeed give the differential equation and its initial condition, resp.

because a series with powers of $V_D(t)$ begins to emerge. Further inserting integrals for U_D reveals that the time-evolution operator can be written as

$$U_D(t,t_0) = 1 + \sum_{n=1}^{\infty} \left(\frac{-i}{\hbar}\right)^n \int_{t'=t_0}^{t} dt' \int_{t''=t_0}^{t'} dt'' \cdots \int_{t^n=t_0}^{t^{n-1}} dt^n V_D(t')V_D(t'')\cdots V_D(t^n). \quad (6.17)$$

This series is called "Dyson series" and the advantage of the Dirac picture is that this structure in powers of the "small" perturbation naturally arises.

For such problems, one might express the state of the system in terms of the unperturbed eigenstates $|n\rangle$:

$$|\phi(t)\rangle = \sum_n c_n(t)|n\rangle \ . \quad (6.18)$$

All states are at $t = t_0$, i.e. solution of the time-independent Schrödinger equation $H_0|n\rangle = E_n|n\rangle$, or might be seen as the time-independent states of the Dirac picture. We now assume that the system starts out in state $|i\rangle$, i.e., $c_n(t_0) = \delta_{n,i}$. After switching on V, the time evolution is given by (6.17) and its matrix elements determine the probability of finding the system in some other state

$$|\phi(t)\rangle_D = U_D(t,t_0)|i(t_0)\rangle_D = U_D(t,t_0)|i\rangle = \underbrace{\sum_n |n\rangle\langle n|}_{=\mathbb{I}} U_D(t,t_0)|i\rangle = \sum_n |n\rangle c_n \ . \quad (6.19)$$

The coefficients $c_n(t)$ can be given in various orders of the perturbation expansion of U_D, see (6.17). In zeroth order $c_n^0(t) = \delta_{i,n}$, in first order, the second line of (6.16) gives

$$U_D^1(t,t_0) = -\frac{i}{\hbar}\int_{t_0}^{t} dt' V_D(t') \ , \quad (6.20)$$

which can be inserted into (6.19) to yield

$$c_n^1(t) = -\frac{i}{\hbar}\int_{t_0}^{t} dt' \langle n|V_D(t')|i\rangle \ . \quad (6.21)$$

When evaluating matrix elements of the Dirac-picture operator V_D with respect to states $|i\rangle$ and $|n\rangle$, use (6.11) to find

$$c_n^1(t) = -\frac{i}{\hbar}\int_{t_0}^{t} dt' \langle n|\mathrm{e}^{\frac{i}{\hbar}H_0(t'-t_0)}V(t')\mathrm{e}^{-\frac{i}{\hbar}H_0(t'-t_0)}|i\rangle = -\frac{i}{\hbar}\int_{t_0}^{t} dt' \mathrm{e}^{i\frac{(E_n-E_i)}{\hbar}(t'-t_0)}\langle n|V(t')|i\rangle \ . \quad (6.22)$$

Similarly, with $t_0 = 0$, second order leads to

$$c_n^2(t) = -\frac{1}{\hbar^2}\sum_m \int_{t_0}^{t} dt' \mathrm{e}^{i\frac{E_n-E_m}{\hbar}t'}\langle n|V(t')|m\rangle \int_{t_0}^{t'} dt'' \mathrm{e}^{i\frac{E_m-E_i}{\hbar}t''}\langle m|V(t'')|i\rangle \ . \quad (6.23)$$

Common approximations consist in stopping the series for $U_D(t,t_0)$ after a few terms, and we are next going to discuss examples.

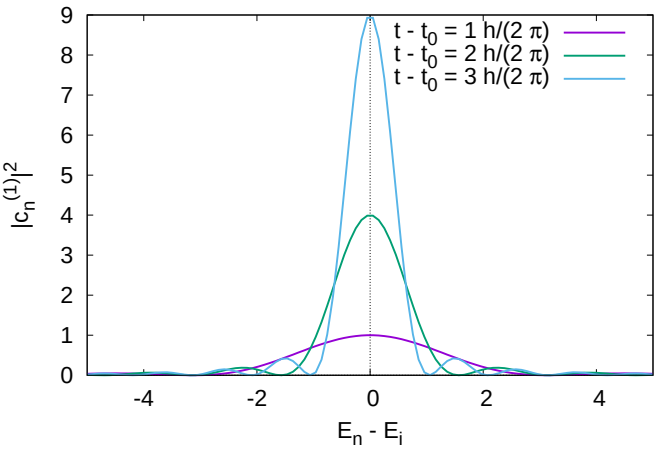


Figure 6.1: Transition probability $|c_n^1(t)|^2$ depending on $E_n - E_i$ for various times $t - t_0$, see (6.25).

6.2.3 Constant perturbation

The time dependence of a "constant" perturbation arises, because it is switched on at time $t_0 = 0$; it remains constant after this time. This may at first seem very similar to a time-*independent* perturbation, but the essential difference is that the system, which is in an unperturbed eigenstate $|i\rangle$ at $t \leq 0$, is not in an eigenstate of the perturbed Hamiltonian and one is interested in the transition elements $c_n(t)$ between the unperturbed initial state and other states $|n\rangle$. In non-degenerate and time-independent perturbation theory, in contrast, one calculates corrections to a given eigenenergy and -state, which the system is assumed to remain in.

With $V(t' > t_0) = \hat{V}$ (note that V is still an operator, not just a number) and $V(t' \leq t_0) = 0$, the first-order correction (6.22) becomes

$$c_n^1(t) = -\frac{i}{\hbar}\langle n|V|i\rangle \int_{t_0}^{t} dt' \mathrm{e}^{i\frac{(E_n-E_i)}{\hbar}(t-t_0)} = -\frac{\langle n|V|i\rangle}{E_n - E_i}\left(\mathrm{e}^{i\frac{(E_n-E_i)}{\hbar}(t-t_0)} - 1\right) . \tag{6.24}$$

The transition probability is

$$\begin{aligned} |c_n^1(t)|^2 &= \frac{\langle n|V|i\rangle^2}{(E_n-E_i)^2}\left(\mathrm{e}^{i\frac{(E_n-E_i)}{\hbar}(t-t_0)} - 1\right)^2 = \frac{2\langle n|V|i\rangle^2}{(E_n-E_i)^2}\left(1 - \cos\frac{E_n-E_i}{\hbar}(t-t_0)\right) = \\ &= \frac{4\langle n|V|i\rangle^2}{(E_n-E_i)^2}\sin^2\left[\frac{(E_n-E_i)}{2\hbar}(t-t_0)\right] , \end{aligned} \tag{6.25}$$

which becomes more and more centered around $E_n - E_i$ as $t - t_0$ becomes longer, see Fig. 6.1. One sees here the connection between energy conservation and time-independence of the Hamiltonian: On short time scales $t \gtrsim t_0$, the Hamiltonian has a pronounced time dependence and the system's energy can change. The Hamiltonian no longer changes on a long time scale, so that energy then has to remain constant.

Using

$$\lim_{a\to\infty} \frac{\sin^2 ax}{x^2} \to \pi a \delta(x) , \tag{6.26}$$

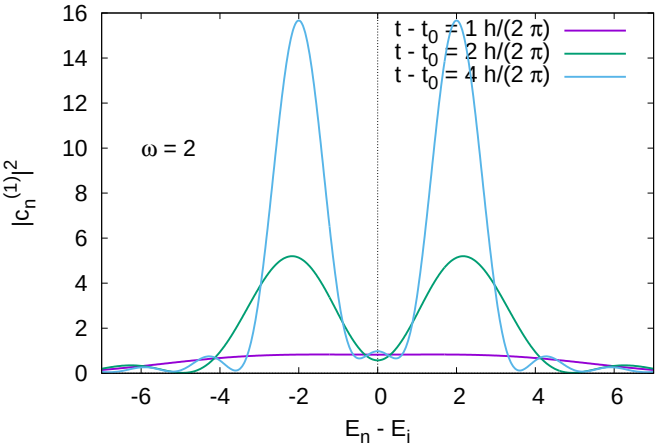


Figure 6.2: Transition probability $|c_n^1(t)|^2$ from (6.27) for $\omega = 2$, depending on $E_n - E_I$ and for various times $t - t_0$.

one finds that the transition rate $\mathrm{d}|c_n^1(t)|^2/\mathrm{d}t$ becomes a constant and involves only states with $E_n = E_i$ (Fermi's golden rule). That said, this limit (long time and $E_n \approx E_i$) is exactly one where perturbation theory cannot be assumed to be valid. Moreover, we can not expect a constant transition rate to hold for all times, as the probability would then keep growing linearly with time. Nevertheless, it turns out that the time can be "long" for the sake of Fermi's golden rule, but still "short" enough for perturbation theory to be valid.

6.2.4 Harmonic perturbation - Interaction with classical radiation

If the interaction after switching on continues to be time-dependent as $V(t' > t_0) = \hat{V}\mathrm{e}^{i\omega t} + \hat{V}^\dagger \mathrm{e}^{-i\omega t}$, the first-order integral is

$$\begin{aligned} c_n^1(t) &= -\frac{i}{\hbar}\langle n|V|i\rangle \int_{t_0}^{t} dt' \mathrm{e}^{i\frac{(E_n-E_i)}{\hbar}(t-t_0)}\mathrm{e}^{i\omega t'} - \frac{i}{\hbar}\langle n|V|i\rangle^* \int_{t_0}^{t} dt' \mathrm{e}^{i\frac{(E_n-E_i)}{\hbar}(t-t_0)}\mathrm{e}^{-i\omega t'} \\ &= -\frac{\langle n|V|i\rangle}{E_n - E_i + \hbar\omega}\left(\mathrm{e}^{i(\frac{(E_n-E_i)}{\hbar}+\omega)(t-t_0)} - 1\right) - \frac{\langle n|V|i\rangle^*}{E_n - E_i - \hbar\omega}\left(\mathrm{e}^{i(\frac{(E_n-E_i)}{\hbar}-\omega)(t-t_0)} - 1\right) . \end{aligned} \tag{6.27}$$

When calculating the transition probability $|c_n^1(t)|$, we can follow the disucssion of Eq. (6.25) and Fig. 6.1: for large $t - t_0$, the square is only large if either $\omega = \frac{(E_n-E_i)}{\hbar}$ or $\omega = -\frac{(E_n-E_i)}{\hbar}$, see Fig. 6.2. The energy of the system can here change, because energy $\hbar\omega$ can be lost (stimulated emission) or gained (stimulated absorption). Spontaneous emission can only be understood by taking into account he quantum nature of light, see Sec. 7.2.3.

Some more information can be obtained from the matrix element $\langle n|V|i\rangle$ if we specify

$$V = -\frac{e}{mc}\hat{\vec{p}}\vec{A} = -\frac{e}{mc}A_0\vec{\epsilon}\vec{p}\left(\mathrm{e}^{i\frac{\omega}{c}\vec{n}\vec{r}-i\omega t} + \mathrm{e}^{-i\frac{\omega}{c}\vec{n}\vec{r}+i\omega t}\right) \tag{6.28}$$

where $\vec{A} = A_0\vec{\epsilon}$ is a vector potential, here chosen to fulfill $\nabla\vec{A} = 0$ (Coulomb gauge) and with polarization ϵ. Operator $\vec{p}$ then acts only on the electronic wave function and commutes

with $\vec{A}$. $|\vec{\epsilon}|^2 = |\vec{n}|^2 = 1$. If we can assume the size of the atom, where the electron is bound, to be much smaller than the wavelength of light, then $\frac{\omega}{c}\vec{n}\vec{r} \approx \frac{\omega}{c}\vec{n}\vec{r_0}$ remains nearly constant over the size of the atom. The resulting complex but constant phase can then be discarded, because only the absolute value enters the transition amplitude $|c_n^1(t)|^2$. This approximation is known as the "electric dipole approximation".

The transition rate in the long-time limit becomes

$$w = \lim_{t\to\infty} \frac{d|c_n^1(t)|^2}{dt} = \frac{2\pi}{\hbar}\frac{e^2}{m^2c^2}|A_0|^2|\vec{\epsilon}\langle n|\vec{p}|i\rangle|^2\delta(E_n - E_i \pm \hbar\omega) \; . \tag{6.29}$$

In order to evaluate this, we select $\epsilon \parallel x$ [2] and make use of $[\hat{x}, H_0] = \frac{i\hbar p_x}{m}$.

$$\langle n|\vec{p}|i\rangle = \frac{-im}{\hbar}\left(\langle n|\hat{x}H_0|i\rangle - \langle n|H_0\hat{x}|i\rangle\right) = \frac{im}{\hbar}(E_n - E_i)\langle n|\hat{x}|i\rangle \tag{6.30}$$

relates the transition rate to the matrix elements of $\vec{x}$. As $\vec{x}$ is odd under parity, one conclusion to be drawn is that transitions are only possible between states with different parity.

[2] General directions for ϵ work analogously.

7 Quantizing Field Theories

Mostly Le Bellac, but beware of typos.

"Second quantization" refers to a formalism that is useful for dealing with many identical particles. It is strongly related to the quantization of classical field theories, where new "particles" arise out of a quantum mechanical treatment of fields. Examples are phonons or photons, the quantized versions of sound and light.

7.1 A Chain of Atoms as Coupled Harmonic Oscillators

Atoms in a solid are on average at their equilibrium position, but will, especially at finite temperature, move around them. The force binding the atoms together can be expanded in a Taylor expansion. the term linear in relative position $x_n - x_{n-1}$ has to vanish, because the atoms would otherwise move to this new equilibrium position. The lowest term is consequently $\propto (x_n - x_{n-1})^2$ and dominates for small movements. (For large movements, effects beyond this term come into play, they are called "anharmonic".) The Hamiltonian of a 1D chain of atoms is then approximately

$$H = \sum_{n=1}^{N} \frac{p_n^2}{2m} + \frac{K}{2} \sum_{n=1}^{N} (q_{n+1} - q_n)^2 \,. \tag{7.1}$$

K is here some "spring constant" coming out of the taylor expansion. As in Sec. 2.1.2, we use periodic boundary conditions, i.e., $p_{n+N} = p_n$ and $q_{n+N} = q_n$.

This Hamiltonian is invariant under translation by multiples of the distance Δ between the atoms, just like the tight-binding case discussed in Sec. 2.1.2. Such a symmetry suggests a conserved quantity and a way to diagonalize the Hamiltonian is to diagonalize an operator related to this symmetry, because it shares an eigensystem with the Hamiltonian. As for the tight-binding Hamiltonian, see Eqs. (2.18) and (2.19), we try this idea by introducing eigenstates of the translation operator:

$$p_k = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} \mathrm{e}^{ikn\Delta} p_n, \quad \text{and} \quad q_k = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} \mathrm{e}^{ikn\Delta} q_n. \tag{7.2}$$

Note that Fourier-transformed versions of both p_n and q_n are used. Formally, this is justified by all of them being canonical, and in this sense "independent", variables in the classical theory. Physically, one should note that the quasi-momentum k is distinct from p_n or p_k. The latter describe the "local" motion of each atom close to its equilibrium positions, while k concerns long-distance properties relating one atom to another. [1] The commutation

[1] In Sec. 2.1.2, we had avoided the issue by simply labeling states by the box m the electron is in. In a full Bloch theory, as in the Kronig-Penney problem from the exercises, one again has both kinds of momentum.

relations (1.47) extended to this case are

$$[p_n, p_m] = 0, \quad [q_n, q_m] = 0, \quad \text{and} \quad [q_n, p_m] = i\hbar\delta_{n,m} \,, \tag{7.3}$$

because operators referring to different atoms commute. They carry over to momentum space, however, with some modification:

$$\begin{aligned}
[p_k, p_{k'}] &= \frac{1}{N}\sum_n\sum_m \mathrm{e}^{ikn\Delta}\mathrm{e}^{ik'm\Delta}\underbrace{[p_n, p_m]}_{=0} = 0 = [q_k, q_{k'}], \\
[q_k, p_{k'}] &= \frac{1}{N}\sum_n\sum_m \mathrm{e}^{ikn\Delta}\mathrm{e}^{ik'm\Delta}\underbrace{[q_n, p_m]}_{=i\hbar\delta_{n,m}} = i\hbar\frac{1}{N}\sum_n \mathrm{e}^{i(k+k')n} = i\hbar\delta_{k,-k'} \,.
\end{aligned} \tag{7.4}$$

(Note the minus sign in $\delta_{k,-k'}$.)

The Hamiltonian in k-space is obtained from the inverse Fourier transform

$$p_n = \frac{1}{\sqrt{N}}\sum_{j_k=-N/2+1}^{N/2} \mathrm{e}^{-i\frac{2\pi j_k}{N}n\Delta}p_n, \quad \text{and} \quad q_n = \frac{1}{\sqrt{N}}\sum_{j_k=-N/2+1}^{N/2} \mathrm{e}^{-i\frac{2\pi j_k}{N}n\Delta}q_n \,. \tag{7.5}$$

Due to periodic boundary conditions, k is restricted to values $j_k\frac{2\pi}{N}$ with integer $j_k = 0, \ldots, N$, which is usually – without any impact on the physics or the mathematics – replaced by the equivalent interval $j_k = -N/2+1, \ldots, N/2$. These are now inserted into (7.1), giving

$$\begin{aligned}
H &= \frac{1}{2mN}\sum_n \left(\sum_k \mathrm{e}^{-ikn\Delta}p_k\right)\left(\sum_{k'} \mathrm{e}^{-ik'n\Delta}p_{k'}\right) \\
&\quad + \frac{K}{2N}\sum_n \left(\sum_k \mathrm{e}^{-ik(n+1)\Delta}q_k - \mathrm{e}^{-ikn\Delta}q_k\right)\left(\sum_{k'} \mathrm{e}^{-ik'(n+1)\Delta}q_{k'} - \mathrm{e}^{-ik'n\Delta}q_{k'}\right) = \\
&= \frac{1}{2m}\sum_k p_k \sum_{k'} p_{k'} \underbrace{\frac{1}{N}\sum_n \mathrm{e}^{-i(k+k')n\Delta}}_{=\delta_{k,-k'}} \\
&\quad + \frac{K}{2}\sum_k q_k(\mathrm{e}^{-ik\Delta}-1)\sum_{k'} q_{k'}(\mathrm{e}^{-ik'\Delta}-1)\underbrace{\frac{1}{N}\sum_n \mathrm{e}^{-i(k+k')n\Delta}}_{=\delta_{k,-k'}} = \\
&= \frac{1}{2m}\sum_k p_k p_{-k} + \frac{K}{2}\sum_k \underbrace{(\mathrm{e}^{-ik\Delta}-1)(\mathrm{e}^{ik\Delta}-1)}_{=2-2\cos k\Delta = 4\sin^2\frac{k\Delta}{2}} q_k q_{-k} = \sum_k \frac{p_k p_{-k}}{2m} + \sum_k \frac{m\omega_k^2}{2}q_k q_{-k}
\end{aligned} \tag{7.6}$$

with $\omega_k = 2\sqrt{\frac{K}{m}}\sin\frac{|k|\Delta}{2}$.

The Hamiltonian is not yet fully decoupled, we now proceed nevertheless with introducing creation and annihilation operators, as in Sec. (1.5.1):

$$a_{-k} = \frac{1}{2}\left(\sqrt{\frac{2m\omega_k}{\hbar}}q_k + i\sqrt{\frac{2}{\hbar m\omega_k}}p_k\right) \quad \text{and} \quad a_k^\dagger = \frac{1}{2}\left(\sqrt{\frac{2m\omega_k}{\hbar}}q_k - i\sqrt{\frac{2}{\hbar m\omega_k}}p_k\right) \,. \tag{7.7}$$

The commutation relations of the new operators are

$$[a_{-k},a_{-k'}]=\frac{1}{4}\left(\frac{2m\omega_k}{\hbar}\underbrace{[q_k,q_{k'}]}_{=0}-\frac{2}{\hbar m\omega_k}\underbrace{[p_k,p_{k'}]}_{=0}+i\frac{2}{\hbar}(\underbrace{[p_k,q_{k'}]}_{-i\hbar\delta_{k,-k'}}+\underbrace{[q_k,p_{k'}]}_{i\hbar\delta_{k,-k'}})\right)=0=[a_k^\dagger,a_{k'}^\dagger]$$
$$[a_k,a_{k'}^\dagger]=\frac{1}{4}\left(\frac{2m\omega_k}{\hbar}\underbrace{[q_{-k},q_{k'}]}_{=0}+\frac{2}{\hbar m\omega_k}\underbrace{[p_{-k},p_{k'}]}_{=0}+i\frac{2}{\hbar}(\underbrace{[p_{-k},q_{k'}]}_{-i\hbar\delta_{k,k'}}-\underbrace{[q_{-k},p_{k'}]}_{i\hbar\delta_{k,k'}})\right)=\delta_{k,k'} \tag{7.8}$$

where we are back to the usual $\delta_{k,k'}$. The Hamiltonian is found by inserting the inverted equations

$$q_k=\sqrt{\frac{\hbar}{2m\omega_k}}\left(a_{-k}+a_k^\dagger\right)\quad\text{and}\quad p_k=-i\sqrt{\frac{\hbar m\omega_k}{2}}\left(a_{-k}-a_k^\dagger\right)\ . \tag{7.9}$$

and becomes

$$H=\frac{-1}{2m}\sum_k\frac{\hbar m\omega_k}{2}\left(a_{-k}-a_k^\dagger\right)\left(a_k-a_{-k}^\dagger\right)+\sum_k\frac{m\omega_k^2}{2}\frac{\hbar}{2m\omega_k}\left(a_{-k}+a_k^\dagger\right)\left(a_k+a_{-k}^\dagger\right)=$$
$$=\sum_k\frac{\hbar\omega_k}{4}(\underbrace{a_{-k}a_{-k}^\dagger}_{=a_{-k}^\dagger a_{-k}+1}+a_k^\dagger a_k+a_{-k}a_{-k}^\dagger+a_k^\dagger a_k)=\sum_k\hbar\omega_k\left(a_k^\dagger a_k+\frac{1}{2}\right)\ . \tag{7.10}$$

It thus becomes the sum over many decoupled harmonic oscillators, where $\omega_{-k}=\omega_k$ has been used.

The eigenstates can be analyzed like those of a single harmonic oscillator, e.g., applying $a_k^\dagger$ to an eigenstate $|\phi\rangle$ (with $H|\phi\rangle=E_\phi|\phi\rangle$) gives another eigenstate

$$Ha_k^\dagger|\phi\rangle=a_k^\dagger H|\phi\rangle+[H,a_k^\dagger]|\phi\rangle=a_k^\dagger E_\phi|\phi\rangle+\sum_{k'}\hbar\omega_{k'}[a_{k'}^\dagger a_{k'},a_k^\dagger]|\phi\rangle=$$
$$=E_\phi a_k^\dagger|\phi\rangle+\sum_{k'}\hbar\omega_{k'}a_{k'}^\dagger\underbrace{[a_{k'},a_k^\dagger]}_{=\delta_{k,k'}}|\phi\rangle=(E_\phi+\hbar\omega_k)a_k^\dagger|\phi\rangle \tag{7.11}$$

whose energy is increased by $\hbar\omega_k$. Similarly, other observations carry over, like the existence of a lowest occupation number $n_k=0$ for each mode. An eigenstate is here characterized by the number of quanta on each mode and is

$$|\phi\rangle=|n_{k_1},n_{k_2},\ldots,n_{k_N}\rangle=\frac{1}{\sqrt{n_{k_1}!}}(a_{k_1}^\dagger)^{n_{k_1}}\otimes\frac{1}{\sqrt{n_{k_2}!}}(a_{k_2}^\dagger)^{n_{k_1}}\otimes\cdots\otimes\frac{1}{\sqrt{n_{k_N}!}}(a_{k_N}^\dagger)^{n_{k_N}}\ , \tag{7.12}$$

as there are N possible modes. Its energy is

$$E_\phi=\sum_k\hbar\omega_k n_k+\underbrace{\frac{\hbar}{2}\sum_k\omega_k}_{=E_0}\ , \tag{7.13}$$

where the second part gives the zero-pont energy E_0. The operator $a_k^\dagger$ still creates an energy quantum as for the harmonic oscillator, but this quantum can also be seen as a "quasi-particle", the phonon. The connection to sound will be made more explicit in the next section.

7.1.1 Connection to Continuous Sound Waves

That the propagation of sound, i.e., waves of moving matter, in an atom chain can be modelled by coupled harmonic oscillators is rather plausible, one sometimes even sees the classical analogue of this as a demonstration experiment. But as we will not have any such classical discretized version for the next section (light), it pays to establish that the classical variant of (7.1) is indeed the Hamiltonian corresponding to discretized sound equations.

To do this, we derive the canonical equations of motion of Hamiltonian classical mechanics:

$$\dot{q_n} = \frac{\partial H}{\partial p_n} = \frac{p_n}{m} \tag{7.14}$$

$$\dot{p_n} = -\frac{\partial H}{\partial q_n} = -K\left[(q_{n+1} - q_n)\cdot(-1) + (q_n - q_{n-1})\right] = K(q_{n+1} + q_{n-1} - 2q_n)\ , \tag{7.15}$$

which can be combined to

$$\ddot{q}_n = \frac{K}{m}(q_{n+1} + q_{n-1} - 2q_n)\ . \tag{7.16}$$

The r.h.s. is very similar to the lowest-order discretization of the second derivative of some function $q(x)$ that coincides with q_n at points $n\Delta$:

$$\frac{\partial^2 q}{\partial x^2} \approx \frac{1}{\Delta^2}(q_{n+1} + q_{n-1} - 2q_n) \tag{7.17}$$

Equation (7.16) can consequently be seen as the discretization of an equation of motion of the continuous function $q(x,t)$,

$$\frac{\partial^2 q}{\partial t^2} - \frac{K\Delta^2}{m}\frac{\partial^2 q}{\partial x^2} = 0, \quad \text{resp.} \quad \frac{\partial^2 q}{\partial t^2} - \underbrace{\frac{Y}{\mu}}_{c_s^2}\frac{\partial^2 q}{\partial x^2} = 0\ . \tag{7.18}$$

In the second version, the microscopic parameters K, m and Δ were replaced by macroscopic $\mu = m/\Delta$ (density = mass per length) and $Y = K\Delta$ (Young's modulus), which are suitable to describe a string. They can be combined to define the velocity of sound, because (7.18) is indeed the wave equation for sound in one dimension. Hamiltonian (7.1) describing coupled harmonic oscillators can thus indeed be taken as a discrete approximation to describe sound, even without knowing that it is a decent model for atoms bound to their equilibrium positions.

One can even go beyond noting that the discrete model should be a good approximation and directly discuss the classical continuous Hamiltonian, obtained as the integral over the Hamiltonian density. A few (mostly technical) differences do arise in this case, i.e. for the (at first classical) Hamiltonian [2]

$$H = \int_0^L dx \left(\frac{p(x)^2}{2\mu} + \frac{\mu c_s^2}{2}\left(\frac{dq(x)}{dx}\right)^2\right)\ . \tag{7.19}$$

[2] Whose canonical equations of motions indeed give directly the continuous wave equation.

One difference is the use of macroscopic parameters, e.g. μ, instead of microscopic ones. Naturally, one should also use the continuous Fourier transform

$$q(k) = \frac{1}{\sqrt{L}} \int_0^L dx\ \mathrm{e}^{ikx} q(x) \quad \text{and} \quad p(k) = \frac{1}{\sqrt{L}} \int_0^L dx\ \mathrm{e}^{ikx} p(x)\ . \tag{7.20}$$

The wave number $k = \frac{2\pi}{L} i_k$ continues to be discrete, because translation by length L should lead back to the original state, i.e., $\mathrm{e}^{ikL} = 1$. In contrast to the coupled atoms, there is no maximal k, and the inverse transform is given by

$$q(x) = \frac{1}{\sqrt{L}} \sum_{i_k=0}^{\infty} \mathrm{e}^{-i\frac{2\pi}{L} i_k x} q(k) \quad \text{and} \quad p(x) = \frac{1}{\sqrt{L}} \sum_{i_k=0}^{\infty} \mathrm{e}^{-i\frac{2\pi}{L} i_k x} p(k)\ . \tag{7.21}$$

Another difference of the continuous case that the derivative dq/dx instead of a difference $q_{n+1} - q_n$ gives a prefactor k^2 (rather than $|1 - \mathrm{e}^{ik\Delta}|^2 = 4 \sin^2 k\Delta/2$) in the Fourier-transformed Hamiltonian

$$H = \sum_k \left[\frac{1}{2\mu} p(k) p(-k) + \frac{\mu c_s^2 k^2}{2} q(k) q(-k) \right]\ . \tag{7.22}$$

The frequency is here linear $\omega_k = c_s |k|$, which is almost the same as before as long as $|k|$ (and ω_k) is small.

With all this in mind, we go once more through some results of the previous section. The classical Hamiltonian (7.22) looks like the one of the discrete case (7.6), and we can accordingly also give an analogous quantum version, by replacing functions $p(k)$ and $q(k)$ with operators that obey the relations (7.3). They can then in turn be expressed in terms of creation and annihilation operators and the solution of this quantum-mechanical version of the continuum Hamiltonian is obtained. By these steps, we "quantize" the previously classical Hamiltonian derived for a continuum Hamiltonian whose equations of motion give the wave equation of sound in one dimension.

The energy of this quantized sound is then given by the sum over harmonic oscillators. As there is no maximal k and as ω_k even grows with k (i.e., definitely does not vanish), the zero-point energy is in fact infinite. This is of no practical relevance if one wants to discuss sound in solids, where a k of a continuous theory can in any case only be considered realistic if $k\Delta$ is not too large. After all, the microscopic structure is expected to matter once the wave oscillates appreciably over its characteristic length scale Δ. One could consequently introduce a cut-off. However, the problem of infinite zero-point energy cannot so easily be argued away in other field theories.

After solving coupled harmonic oscillators, one can of course go back and express $\hat{q}_n$ in terms of $\hat{q}_k$, which is in turn given by a_{-k} and $a_k^\dagger$. After noting the similarity of this situation to sound, one can do the same and insert $a_{-k}^\dagger$ and a_k via $q(k)$ into the continuous function $q(x)$, which then becomes operator-valued:

$$\begin{aligned} \hat{q}(x) &= \frac{1}{\sqrt{L}} \sum_k \mathrm{e}^{-ikx} \hat{q}(k) = \sqrt{\frac{\hbar}{2\mu L}} \sum_k \sqrt{\frac{1}{\omega_k}} \left(\mathrm{e}^{-ikx} a_{-k} + \mathrm{e}^{-ikx} a_k^\dagger \right) = \\ &= \sqrt{\frac{\hbar}{2\mu L}} \sum_k \sqrt{\frac{1}{\omega_k}} \left(\mathrm{e}^{ikx} a_k + \mathrm{e}^{-ikx} a_k^\dagger \right)\ . \end{aligned} \tag{7.23}$$

The last expression was here obtained by rearranging terms with k and $-k$. The classical field $q(x)$ has here been quantized, note that x itself is no longer a quantum mechanical operator (as it was for the single harmonic oscillator), but a parameter, on which $\hat{q}$ depends. One can also give the quantized version of the canonically conjugate momentum field

$$\hat{p}(x) = -i\sqrt{\frac{\hbar\mu}{2L}}\sum_k \sqrt{\omega_k}\left(\mathrm{e}^{ikx}a_k - \mathrm{e}^{-ikx}a_k^\dagger\right) \ . \tag{7.24}$$

One can check that

$$\begin{aligned}[\hat{q}(x),\hat{p}(x')] &= -i\sqrt{\frac{\hbar}{2\mu}}\sqrt{\frac{\hbar\mu}{2}}\frac{1}{L}\sum_k\sum_{k'}\sqrt{\frac{\omega_{k'}}{\omega_k}}\left(\mathrm{e}^{-ikx}\mathrm{e}^{ik'x'}\underbrace{[a_k^\dagger,a_{k'}]}_{=-\delta_{k,k'}} - \mathrm{e}^{ikx}\mathrm{e}^{-ik'x'}\underbrace{[a_k,a_{k'}^\dagger]}_{=\delta_{k,k'}}\right) \\ &= \frac{i\hbar}{L}\sum_k \mathrm{e}^{ik(x-x')} = i\hbar\delta(x-x')\end{aligned} \tag{7.25}$$

as long as the momentum-dependent operators satisfy (7.3).

As the sound equation contains derivatives in time and space on the same footing, we go to the Heisenberg picture and derive a time-dependent operator version $\hat{q}(x,t)$. The equation for the time evolution of the annihilation operator was given in Eq. (6.8), it and its analogue for the creation operator can be integrated to

$$a_{k,H}(t) = a_k(t) = \mathrm{e}^{-i\omega_k t}a_k \quad \text{and} \quad a_{k,H}^\dagger(t) = a_k^\dagger(t) = \mathrm{e}^{i\omega_k t}a_k^\dagger \ , \tag{7.26}$$

where the subscript H is usually omitted. The time-dependent operators are now inserted in the continuum versions of (7.9) and Fourier-transformed back into position space to yield

$$\begin{aligned}\hat{q}(x,t) &= \sum_k \mathrm{e}^{-ikx}\hat{q}(k,t) = \sqrt{\frac{\hbar}{2\mu L}}\sum_k\sqrt{\frac{1}{\omega_k}}\left(\mathrm{e}^{-ikx-i\omega_k t}a_{-k} + \mathrm{e}^{-ikx+i\omega_k t}a_k^\dagger\right) \\ &= \sqrt{\frac{\hbar}{2\mu L}}\sum_k\sqrt{\frac{1}{\omega_k}}\left(\mathrm{e}^{ikx-i\omega_k t}a_k + \mathrm{e}^{-ikx+i\omega_k t}a_k^\dagger\right) \ .\end{aligned} \tag{7.27}$$

If one had set out to derive a quantum theory of sound, this might – with hindsight – have worked in the following way:

1. Start from the continuous wave equation (7.18).
2. Obtain the corresponding Hamiltonian density in position space, i.e., depending on x and the Hamiltonian (7.19) as the integral.
3. Fourier transform it into k-space, giving (7.22)
4. Notice that this looks like coupled harmonic oscillators, which we understand well.
5. Accordingly quantize this Hamiltonian, i.e., $p(k)$ and $q(k)$ become operators (and should perhaps be written $\hat{p}(k)$ and $\hat{q}(k)$) with the canonical commutation relations (7.3).
6. Solve the coupled harmonic oscillators as done above by introducing $a_k^\dagger$ and a_k.
7. Express the original field(s) $q(x,t)$ (and $p(x,t)$) in terms of $a_k^\dagger$ and a_k, see (7.27).

7.2 Quantization of the Electromagnetic Field

When trying to find a quantum theory for classical electromagnetism, there is no plausible discrete classical version similar to the chain of oscillating atoms. The wave equation is analogous to that of sound, however, so we can hope that the procedure outlined above still works.

In Si units, the Maxwell equations are

$$\nabla B = 0 \qquad \nabla \times E = -\frac{\partial B}{\partial t} \tag{7.28}$$

$$\nabla E = \frac{\rho}{e} \qquad c^2 \nabla \times B = \frac{\partial E}{\partial t} + \frac{j}{e} \tag{7.29}$$

where ρ and j are electric charge and current, E and B are electric and magnetic field and c is the speed of light. We now set ρ and j to zero. As E and B depend on each other, it is not clear we can use them immediately as the canonical dynamic variables, which should be independent. Instead, we go to the vector and scalar potentials A and ϕ, where

$$E = \nabla\phi - \frac{\partial A}{\partial t}, \quad B = \nabla \times A \tag{7.30}$$

These are not unique, we choose Coulomb gauge, where $\nabla A = 0$. One then finds from (7.29) that

$$\nabla E = \nabla \left(\nabla\phi - \frac{\partial A}{\partial t} \right) = \Delta\phi + \frac{\partial}{\partial t} \nabla A = \Delta\phi = -\frac{\rho}{e} \,, \tag{7.31}$$

which vanishes as $\rho = 0$. (Moreover, we can deduce $\phi(r,t) = \int d^3r' \frac{\rho(r',t)}{|r-r'|}$.)

The time derivative of E is then given on one hand by $\frac{\partial E}{\partial t} = -\frac{\partial^2 A}{dt^2}$ and on the other by

$$\frac{\partial E}{\partial t} = c^2 \nabla \times (\nabla \times A) = c^2 \nabla \cdot \underbrace{(\nabla A)}_{=0} - c^2 \Delta A \tag{7.32}$$

and the wave equation

$$\frac{\partial^2 A}{\partial t^2} - c^2 \Delta A = 0 \tag{7.33}$$

results. We also know the classical energy density $E^2 + c^2 B^2$ that determines the classical Hamiltonian

$$H = \int h(r) \mathrm{d}^3 r = \frac{\epsilon_0}{2} \int \mathrm{d}^3 r \left(E(r)^2 + c^2 B(r)^2 \right) \,. \tag{7.34}$$

One can check that this energy density indeed gives the wave equation from its equations of motion by working out the functional derivatives with respect to the canonical variables – which we still need to identify. E is a time derivative of A and B involves its spatial derivatives, which is a similar structure to p and $\frac{\partial q}{\partial x}$ in Eq. (7.19). Tentatively, one might thus assign one dynamic variable ("p") to be proportional to E and the other ("q") to A.

One would then Fourier transform the Hamiltonian by writing it it terms of the canonical variables in k-space, where a few technical differences to the 1D case arise due to the curl in B. The Fourier transformed Hamiltonian in the proper variables is then quantized. All these steps are detailed in the following subsection.

7.2.1 Transformation and quantization of the Hamiltonian

In this subsection, the details alluded to above are given. From the derivative

$$\frac{\partial H}{\partial E(r)} = \epsilon_0 E(r) = -\epsilon_0 \frac{\partial A(r)}{\partial t} \,, \tag{7.35}$$

one can guess that $-\epsilon E(r)$ would make a good candidate for "p", because

$$\frac{\partial H}{\partial(-\epsilon_o E(r))} = \frac{\partial A(r)}{\partial t} \,, \tag{7.36}$$

The role of q should then be taken by $A(r)$, which needs to be checked via the second equation of motion. Due to the curl, we need to be more careful here, the explicit functional derivative with respect to component A_x is

$$\frac{\delta H}{\delta A_x} = \Big[\frac{d}{d\varepsilon}\int h(E, A+\varepsilon\phi(r)e_x, \underbrace{\nabla\times(A+\varepsilon\phi(r)e_x)}_{=\nabla\times A+\varepsilon\nabla\times\phi(r)e_x})d^3r\Big]_{\varepsilon=0} \,, \tag{7.37}$$

where $\phi(r)$ is a test function and e_x the unit vector in x direction.

$$\frac{\delta H}{\delta A_x} = \Big[\frac{d}{d\varepsilon}\int d^3r \ \underbrace{\frac{\partial h}{\partial(\nabla\times A)}}_{=\epsilon_0 c^2 B}\cdot\varepsilon\nabla\times\phi(r)e_x\Big]_{\varepsilon=0} = \epsilon_0 c^2\int d^3r \ B\cdot\nabla\times\phi(r)e_x =$$

$$= \epsilon_0 c^2\int d^3r \ [B_y\partial_z\phi(r) - B_z\partial_y\phi(r)]$$

As $\phi(r)$ is a test function that we want in the end to just stand at the end of the integral, we need to shift the derivatives from $\phi(r)$ to B. This is done via $\partial_\alpha(B_\beta\phi(r)) = \phi(r)\partial_\alpha B_\beta + B_\beta\partial_\alpha\phi(r)$:

$$\frac{\delta H}{\delta A_x} = \epsilon_0 c^2\int d^3r \ \big[\partial_z(B_y\phi(r)) - \partial_y(B_z\phi(r))\big] - \epsilon_0 c^2\int d^3r \ \big[(\partial_z B_y) - (\partial_y B_z)\big]\phi(r) \,.$$

In the second integral, the derivatives do not apply to $\phi(r)$ any longer. The first integral consists of total derivatives, the integral thus only depends on the boundary and is given by $B_{y/z}\phi(r)$. Since $\phi(r)$ is a test function, it is assumed to vanish sufficiently strongly at the boundary to make this first integral vanish. We are then left with

$$\frac{\delta H}{\delta A_x} = -\epsilon_0 c^2\int d^3r \ \big[(\partial_z B_y) - (\partial_y B_z)\big]\phi(r) = \epsilon_0 c^2\int d^3r \ (\nabla\times B)\cdot e_x\phi(r) \,, \tag{7.38}$$

because $\partial_y B_z - \partial_z B_y$ is the x-component of the curl of B. In a vector notation, one then finds

$$\frac{\partial H}{\partial A(r)} = \epsilon_0 c^2\nabla\times B = \epsilon_0 c^2\nabla\times(\nabla\times A) = -\epsilon_0 c^2\Delta A \,. \tag{7.39}$$

According to Hamiltonian mechanics

$$\frac{\partial H}{\partial A(r)} = -\frac{\partial(-\epsilon_0 E)}{\partial t} \,, \tag{7.40}$$

and hence

$$-\epsilon_0 c^2 \Delta A = -\epsilon_0 \frac{\partial^2 A}{\partial t^2} \ , \tag{7.41}$$

which reproduces the wave equation. Hamiltonian (7.34) with canonical variables $A(r)$ and $-\epsilon_0 E(r)$ (in the roles of q and p, respectively) thus describe (classical) light.

The Fourier transform of the Hamiltonian works as follows:

$$\begin{aligned} H &= \int d^3r \Big[\frac{(\epsilon_0 E)^2}{2\epsilon_0} + \frac{\epsilon_0 c^2}{2} (\nabla \times A)^2 \Big] = \frac{1}{L^3} \sum_k \sum_{k'} \int d^3r \frac{(\epsilon_0 E(k) \mathrm{e}^{-ikr})(\epsilon_0 E(k') \mathrm{e}^{-ik'r})}{2\epsilon_0} \\ &\quad + \frac{1}{L^3} \sum_k \sum_{k'} \int d^3r \frac{\epsilon_0 c^2}{2} \underbrace{(\nabla \times e^{-ikr} A(k))}_{=(\nabla e^{-ikr}) \times A(k)} (\nabla \times e^{-ik'r} A(k')) = \\ &= \frac{1}{2\epsilon_0} \sum_k \sum_{k'} (\epsilon_0 E(k))(\epsilon_0 E(k')) \underbrace{\frac{\int d^3r \mathrm{e}^{-i(k+k')r}}{L^3}}_{=\delta_{k,-k'}} \\ &\quad + \frac{\epsilon_0 c^2}{2L^3} \sum_k \sum_{k'} \int d^3r \ (e^{-ikr}(-ik) \times A(k))(e^{-ik'r}(-ik') \times A(k')) = \\ &= \frac{1}{2\epsilon_0} \sum_k (\epsilon_0 E(k))(\epsilon_0 E(-k)) + \frac{\epsilon_0 c^2}{2} \sum_k -(k \times A(k))(-k \times A(-k)) \ . \end{aligned} \tag{7.42}$$

The first term looks like the p-term of sound, except that $-\epsilon_0 \vec{E}(\vec{k})$ are vectors and the dot product thus implies a sum of the three components. In the second term, only components of $A(k)$ orthogonal to k enter – in fact,

$$0 = \nabla A(r) = \frac{1}{\sqrt{L^3}} \sum_k \nabla \mathrm{e}^{-ikr} A(k) = \frac{-i}{\sqrt{L^3}} \sum_k \mathrm{e}^{-ikr} k \cdot A(k) \tag{7.43}$$

implies $k \cdot A(k) = 0$, as it has to vanish for arbitrary r.

Vector $A(k)$ can be written as the sum over its two components orthogonal to k, $\vec{A}(k) = A_1(k)\vec{e}_1^*(k) + A_2(k)\vec{e}_2^*(k)$, with $\vec{e}_1^*(k) \cdot \vec{e}_2(k) = \vec{e}_1(k) \cdot \vec{e}_2^*(k) = 0$ and $\vec{e}_s^*(k) \cdot k = \vec{e}_s(k) \cdot k = 0$. ($A_1(k)$ and $A_2(k)$ are no longer vectors.) The unit vectors can be complex, e.g., for circular polarization. The use of the complex conjugate in this definition ensures a "conventional" notation later, but is not necessary. The cross products then become

$$k \times A(k) = k \times A_1(k)\vec{e}_1^*(k) + k \times A_2(k)\vec{e}_2^*(k) = |k| A_1(k)\vec{e}_2^*(k) - |k| A_2(k)\vec{e}_1^*(k) \tag{7.44}$$

With the choice $\vec{e}_s^*(-k) = \vec{e}_s(k)$, on finds

$$(k \times A(k))(k \times A(-k)) = |k|^2 \big(A_1(k)A_1(-k) + A_2(k)A_2(-k) \big) = \sum_{s=1}^{2} |k|^2 A_s(k) A_s(-k) \ . \tag{7.45}$$

The k-dependent classical Hamiltonian is then

$$H = \frac{1}{2\epsilon_0} \sum_k \sum_{s=1}^{2} (\epsilon_0 E_s(k))(\epsilon_0 E_s(-k)) + \frac{\epsilon_0}{2} \sum_k \sum_{s=1}^{2} \underbrace{c^2 |k|^2}_{=\omega_k^2} A_s(k) A_s(-k) \ , \tag{7.46}$$

where the analogy to (7.22) is obvious and ϵ_0 takes the role μ.

This Hamiltonian is now quantized by introducing operators

$$\hat{A}_s(k) = \sqrt{\frac{\hbar}{2\epsilon_0\omega_k}}\left(a_{-k,s} + a^\dagger_{k,s}\right) \quad \text{and} \quad -\epsilon_0\hat{E}_s(k) = -i\sqrt{\frac{\hbar\epsilon_0\omega_k}{2}}\left(a_{-k,s} - a^\dagger_{k,s}\right) \tag{7.47}$$

For these operators, the commutation relations

$$[\hat{E}_s(k), \hat{E}_{s'}(k')] = [\hat{A}_s(k), \hat{A}_{s'}(k')] = 0 \tag{7.48}$$

$$[\hat{A}_{s'}(k'), -\epsilon_0\hat{E}_s(k)] = i\hbar\delta_{s,s'}\delta_{k,-k'} \tag{7.49}$$

are postulated, which is equivalent to requiring

$$[a_{k,s}, a_{k',s'}] = [a^\dagger_{k,s}, a^\dagger_{k',s'}] = 0 \tag{7.50}$$

$$[a_{k,s}, a^\dagger_{k',s'}] = \delta_{s,s'}\delta_{k,k'} \ . \tag{7.51}$$

7.2.2 Quantum Theory of Light

From analogy with sound as well as from the preceding subsection, it follows that the solutions for quantized sound can essentially be copied. $\vec{A}$ is a vector, where Coulomb gauge imposes $\vec{k}\cdot\vec{A} = 0$. The two remaining "transverse" modes double the modes of the harmonic oscillator, so that the quantized Hamiltonian expressed with $a^\dagger_{\vec{k},s}$ and $a_{\vec{k},s}$ is

$$H = \sum_{\vec{k}}\sum_{s=1}^{2}\hbar\omega_{\vec{k}}\left(a^\dagger_{\vec{k},s}a_{\vec{k},s} + \frac{1}{2}\right) \ , \tag{7.52}$$

with frequency $\omega_{\vec{k}} = c|k|$. Operator $a^\dagger_{\vec{k},s}$ creates an energy quantum with momentum $\vec{k}$ and polarization s, a "photon". The eigenstates are obtained by populating the modes $(\vec{k}, s)$ with $n_{\vec{k},s}$ photons, as in (7.12); the corresponding eigenenergy E is given by the number of photons in each mode $\vec{k}$:

$$H|n_{\vec{k}_1,s_1}, n_{\vec{k}_1,s_2}, n_{\vec{k}_2,s_1}, \dots\rangle = \underbrace{\sum_{\vec{k},s}\hbar\omega_{\vec{k}}\left(n_{\vec{k},s} + \frac{1}{2}\right)}_{E}|n_{\vec{k}_1,s_1}, n_{\vec{k}_1,s_2}, n_{\vec{k}_2,s_1}, \dots\rangle \ . \tag{7.53}$$

The modes $\vec{k}$ are discrete as long as L is finite, but their size is not bounded and there is no good reason for a cut-off. The zero-point energy is consequently infinite. In most cases, this is not a problem, because only energy *differences* are relevant, but the issue is debated as potentially relevant in cosmology.

In addition to replacing c_s by c, $\mathrm{e}^{i\vec{k}x}$ is replaced by $\mathrm{e}^{i\vec{k}\vec{r}}$, L by L^3 and μ by ϵ_0, giving the Schrödinger operator describing the quantized vector field as

$$\hat{A}(\vec{r}, t) = \sqrt{\frac{\hbar}{2\epsilon_0 L^3}}\sum_{\vec{k}}\sum_{s=1}^{2}\sqrt{\frac{1}{\omega_{\vec{k}}}}\left(\mathrm{e}^{i\vec{k}\vec{r}}\vec{e}_s a_{\vec{k},s} + \mathrm{e}^{-i\vec{k}\vec{r}}\vec{e}^*_s a^\dagger_{\vec{k},s}\right) \ . \tag{7.54}$$

and the time-dependent Heisenberg operator as

$$\hat{A}(\vec{r},t) = \sqrt{\frac{\hbar}{2\epsilon_0 L^3}} \sum_{\vec{k}} \sum_{s=1}^{2} \sqrt{\frac{1}{\omega_{\vec{k}}}} \left(\mathrm{e}^{i(\vec{k}\vec{r}-\omega_{\vec{k}}t)} \vec{e}_s a_{\vec{k},s} + \mathrm{e}^{-i(\vec{k}\vec{r}-\omega_{\vec{k}}t)} \vec{e}_s^{\,*} a^\dagger_{\vec{k},s} \right) . \tag{7.55}$$

in analogy to (7.27). One can check that $E(x,t)$ obtained from the canonical conjugate $-\epsilon\hat{E}$ is indeed the time derivative, namely

$$\hat{E}(\vec{r},t) = i\sqrt{\frac{\hbar}{2\epsilon_0 L^3}} \sum_{\vec{k}} \sum_{s=1}^{2} \sqrt{\omega_{\vec{k}}} \left(\mathrm{e}^{i(\vec{k}\vec{r}-\omega_{\vec{k}}t)} \vec{e}_s a_{\vec{k},s} - \mathrm{e}^{-i(\vec{k}\vec{r}-\omega_{\vec{k}}t)} \vec{e}_s^{\,*} a^\dagger_{\vec{k},s} \right) . \tag{7.56}$$

The quantized magnetic field is obtained as

$$\begin{aligned}
\hat{B}(x,t) = \nabla \times \hat{A}(r,t) &= \sqrt{\frac{\hbar}{2\epsilon_0 L^3}} \sum_{\vec{k}} \sum_{s=1}^{2} \sqrt{\frac{1}{\omega_{\vec{k}}}} \left(\nabla \times (\mathrm{e}^{i(\vec{k}\vec{r}-\omega_{\vec{k}}t)} \vec{e}_s) a_{\vec{k},s} + \nabla \times (\mathrm{e}^{-i(\vec{k}\vec{r}-\omega_{\vec{k}}t)} \vec{e}_s^{\,*}) a^\dagger_{\vec{k},s} \right) = \\
&= \sqrt{\frac{\hbar}{2\epsilon_0 L^3}} \sum_{\vec{k}} \sum_{s=1}^{2} \sqrt{\frac{1}{\omega_{\vec{k}}}} \left(i\vec{k} \times \vec{e}_s \mathrm{e}^{i(\vec{k}\vec{r}-\omega_{\vec{k}}t)} a_{\vec{k},s} - i\vec{k} \times \vec{e}_s^{\,*} \mathrm{e}^{-i(\vec{k}\vec{r}-\omega_{\vec{k}}t)} a^\dagger_{\vec{k},s} \right) = \\
&= i\sqrt{\frac{\hbar}{2\epsilon_0 L^3}} \sum_{\vec{k}} \sum_{s=1}^{2} \frac{\sqrt{\omega_{\vec{k}}}}{c} \left(\frac{\vec{k}}{|\vec{k}|} \times \vec{e}_s \mathrm{e}^{i(\vec{k}\vec{r}-\omega_{\vec{k}}t)} a_{\vec{k},s} - \frac{\vec{k}}{|\vec{k}|} \times \vec{e}_s^{\,*} \mathrm{e}^{-i(\vec{k}\vec{r}-\omega_{\vec{k}}t)} a^\dagger_{\vec{k},s} \right)
\end{aligned} \tag{7.57}$$

The quantization program outlined at the end of Sec. 7.1.1 works here, as well. In fact, it is not necessary to perform all the steps: Once the canonical variables have been identified (which may be an effort for a field thoery), they can be quantized. Whether one wants to do this in position or momentum space is at most a matter of taste.

7.2.3 Interaction of an electron with quantum-mechanical light

In class, we had used time-dependent perturbation theory to discuss the electronic transitions in an oscillating magnetic field, a classical description of light, see Sec. 6.2.4. The result was that a changing electromagnetic field can induce transitions between states of different energy, stimulated emission/absorption. We are now in a position to extend the discussion to a quantum mechanical treatment of light.

The Hamiltonian might be written as

$$H = H_{\text{electron}} + H_{\text{light}} + H_{\text{interaction}} \tag{7.58}$$

where $H_0 = H_{\text{electron}} + H_{\text{light}}$ is the unperturbed Hamiltonian of one electron and of light that do not interact. The interaction is captured in $H_{\text{interaction}}$, which is assumed to be a "small" perturbation V. Eigenstates of H_0, i.e., describing the noninteracting subsystems, can be written as tensor products $|\phi^0_n\rangle = |\phi^0_{el,n}\rangle \otimes |\phi^0_{e.-m.,n}\rangle$ consisting of one part for the electron and one for the radiation. The unperturbed states of the electron might, for example, be bound states in an atomic potential; unperturbed states of the electromagnetic field are given by the number $n_{\vec{k},s}$ of photons in each mode $(\vec{k},s)$. The total energy of an

unperturbed state is the sums of the energy of the atomic state and energy (7.53) of the radiation field.

The interaction Hamiltonian is again described with "minimal coupling", i.e., by replacing $\hat{p}$ of the electron by $\hat{p}+q_e\hat{A}$, where q_e is the charge of the electron. The quantum-mechanical field operator is given in (7.54). The term q_eA^2 only affects the electromagnetic field and its impact on the electron (via the light) is to be small. This gives

$$V = H_{\text{interaction}} = \frac{2q_e\hat{p}\hat{A}}{2m} = \frac{q_e}{m_e}\hat{p}\hat{A} \,. \tag{7.59}$$

p and the creation/annihilation operators $a^\dagger/a$ in A commute in any case, because they act on different subsystems; p and the "prefactor" $\mathrm{e}^{i\vec{k}\vec{r}}$ commute here, because of the use of Coulomb gauge, where $pA \propto \nabla A = k \cdot A = 0$. The interaction is now treated in time-dependent perturbation theory, where the time-dependence is the "switching on" of the interaction, e.g., by bringing the atom with the electron close enough to a radiation field. However, the perturbation is then constant in time, as in Sec. 6.2.3 rather than oscillating as in Sec. 6.2.4.

The first-order result of time-dependent perturbation theory is

$$c_n^1(t) = -\frac{i}{\hbar}\langle n|V|i\rangle \int_{t_0}^{t} dt' \mathrm{e}^{i\frac{(E_n-E_i)}{\hbar}(t-t_0)} = -\langle n|V|i\rangle \frac{\mathrm{e}^{i\frac{(E_n-E_i)}{\hbar}(t-t_0)}-1}{E_n-E_i} \,. \tag{7.60}$$

The energy-dependence only gives a significant contribution for long t if $E_n - E_i \approx 0$, and we need to evaluate the matrix elements $\langle n|V|i\rangle$.

The matrix element contains a product of two operators, each of which acts on a different subsystem, $\hat{p}$ on the electronic state in the atom and $\hat{A}$ on the radiation field:

$$\frac{q_e}{m_e}\langle n|\hat{p}\otimes\hat{A}|i\rangle = \frac{q_e}{m_e}\langle\phi^0_{el,n}|\hat{p}|\phi^0_{el,i}\rangle\langle\phi^0_{e.-m.,n}|\hat{A}|\phi^0_{e.-m.,i}\rangle \,. \tag{7.61}$$

The factor from the electronic wave functions follows the discussion of interaction with classical radiation as discussed in Sec. 6.2.4. Concerning the now quantized vector potential $\hat{A}(\vec{r},t)$, see (7.54), the polarization $\vec{\epsilon}_s$ is again included in the electronic part and the spatial phase $\mathrm{e}^{i\vec{k}\vec{r}} \approx \mathrm{e}^{i\vec{k}\vec{r}_0} = \mathrm{e}^{i\omega_{\vec{k}}r_0/c}$ is again assumed to be constant due to the atom's small size compared to the wave length of light, both as in Sec. 6.2.4.

Apart from constant prefactors, the transition rate in the large-t limit is then

$$\Big|\sum_{s=1}^{2}\vec{e}_s\langle\phi^0_{el,n}|\hat{p}|\phi^0_{el,i}\rangle\sum_{\vec{k}}\sqrt{\frac{1}{\omega_{\vec{k}}}}\langle\phi^0_{e.-m.,n}|a_{\vec{k},s}|\phi^0_{e.-m.,i}\rangle + \tag{7.62}$$
$$\sum_{s=1}^{2}\vec{e}^{\,*}_s\langle\phi^0_{el,n}|\hat{p}|\phi^0_{el,i}\rangle\sum_{\vec{k}}\sqrt{\frac{1}{\omega_{\vec{k}}}}\langle\phi^0_{e.-m.,n}|a^\dagger_{\vec{k},s}|\phi^0_{e.-m.,i}\rangle\Big|^2 \cdot \delta(E_{el,n}+E_{e.-m.,n}-E_{el,i}-E_{e.-m.,i}) \,.$$

Due to energy conservation, energy differences $E_{el,n} - E_{el,i}$ of the electron have to be made up by those of the photon field. As operators a ($a^\dagger$) annihilate (create) one photon, the difference between the energy (7.53) of the two states must be $E_{e.-m.,n} - E_{e.-m.,i} = -\hbar\omega_{\vec{k}}$ ($E_{e.-m.,n} - E_{e.-m.,i} = \hbar\omega_{\vec{k}}$), implying $E_{el,n} - E_{el,i} = \pm\hbar\omega_{\vec{k}}$.

As allowed $\pm\hbar\omega_{\vec{k}}$ are well spaced in energy for discrete atomic levels, we can discuss each term in (7.62) separately. For the transition rate, factors $\propto |\langle\phi^0_{e.-m.,n}|a_{\vec{k},s}|\phi^0_{e.-m.,i}\rangle|^2 = n_{\vec{k},s}$ and $\propto |\langle\phi^0_{e.-m.,n}|a^\dagger_{\vec{k},s}|\phi^0_{e.-m.,i}\rangle|^2 = n_{\vec{k},s}+1$ arise, i.e., absorption an emission rates are higher if there are more photons. As this is roughly $\propto |A_0|^2$ in the case of large n, the results agrees with those for stimulated absorption and emission due to classical light.

An important qualitative difference arises, however, for emission and the limit of small $|A_0|$: The factor $|\langle\phi^0_{e.-m.,n}|a^\dagger_{\vec{k},s}|\phi^0_{e.-m.,i}\rangle|^2 = n_{\vec{k},s}+1$ is also finite for $n_{\vec{k},s}=0$, i.e., if there are no photons present in the initial state. This corresponds to *spontaneous emission* and cannot be captured by a classical treatment. Unless the electron wave function is such that $\langle\phi^0_{el,n}|\hat{p}|\phi^0_{el,i}\rangle = 0$ for all $E_{el,n} < E_{el,i}$, an excited state will thus at some point emit a photon and go into a lower-energy state, giving it a finite life time.

8 "Second quantization": Creation and Annihilation operators for Bosons and Fermions

Partly Sakurai II, partly Nolting (Many-body)

In treating the harmonic oscillator, the creation/annihilation operators were useful to go from one eigenstate to others. While they can be seen the create/annihilate "energy quanta", this is just one way to see them. (An analogous interpretation is after all not usually applied to ladder operators for angular momentum.) In the case of coupled harmonic oscillators, energy quanta were identified as phonons, and for the electro-magnetic field, we finally found photons with momentum and spin - quite particle-like entities. In this section, we are going to see that we can turn this around and describe theories formulated for particle with creation/annihilation operators as well. This formalism is particularly suitable to problems involving many identical particles.

8.1 Identical Particles

Even in classical statistical physics, one only deals with averages and does not aim to describe all individual particles of a large ensemble, because this would not be practically feasible. Quantum mechanics, however, takes this one step further and states that identical particles (e.g. several electrons) can on principle not be distinguished and that this has to be taken into account.

If we have two identical particles in two different states a and b and exchange them, no observable should be affected: This is the the requirement that they be indistinguishable. However, the total wave function of this exchanged state can still differ from the original one by a phase $\mathrm{e}^{i\phi_{ab}}$. We are going to see that this phase is ± 1 (except in the very special case of 2D anyons) and does not depend on states a and b.

8.1.1 Anyons

In 3D, exchanging the particles twice, so that in the end particle 1 is back in state a and 2 in state b cannot lead to a phase depending only in a and b. The formal proof involves homotopy analysis of the group of permutations, but one can also give an intuitive picture:

1. State a is on top and state b below it, i.e., they are both given by position eigenstates on the z-axis, with a at height a and b at 0.
2. An exchange is performed by moving particle 1 from a to a point on the x-axis and further down to $-a$. Particle 2 is then moved up to $+a$ and particle 1 finally goes to 0, i.e., position b.

3. The second exchange can be performed by moving particle 1 from position b (at 0) to a point on the negative x-axis and then up to $2a$. Particle 2 is then lowered to 0 (bringing it back to its original position) and particle 1 finishes by going down to a again.
4. This double exchange should lead to the same wave function as the original, but possibly with an additional phase $\mathrm{e}^{i\phi_{ab}}$.
5. However, the second exchange could also have been performed as follows: Move 1 from 0 down to $-a$, move 2 from a to 0, move 1 from $-a$ to the point along the x-axis and up to a, i.e., by reversing the steps of the original swap.
6. Reversing all the steps of the original swap should also undo any phases picked up, so this second way should not lead to any phase.
7. Finally, the two ways to perform the second swap can be continuously connected: Effectively, particle 1 can go around particle 2 (as in the first example) or not (as in the second) and in 3D, the two ways can be distorted into each other via the y-direction.
8. Any phase would consequently have to interpolate from $\theta_{ab} = 0$ to finite θ_{ab} depending on the path. The wave function would thus have to encode all the detailed history of the swaps as well, i.e., simple permutations would not do it. There is no evidence of this.
9. Consequently the phase of a double swap has to be 1.

An exception to this argument can be found if the particles live in 2D, e.g. in the x-z plane. Here, the two ways to swap the particles back cannot be smoothly distorted into each other: Either particle 1 goes once around particle 2 or it goes down and up again on the same side. (The y direction does not exist.) If it goes around, it may pick up a phase that does not depend on the details of the path but only on the winding number. It turns out that quasiparticles with such properties can indeed arise in fractional quantum Hall states, which arise in 2D semiconductors. As they can have any phase upon particle exchange rather than just ± 1, they are called anyons. The search for a kind of anyons whose permutations do not commute is currently a hot topic, because such wave functions might have a potential for quantum computing.

8.1.2 Bosons and Fermions

This part from Le Bellac

Let $P_{1,2}$ denote the operator that switches particles 1 and 2. As it is not supposed to have any impact on observable features, e.g., on scalar products, it has to be unitary $P_{1,2}^\dagger = P_{1,2}^{-1}$. Moreover, applying it twice has to give the identity operator (except for anyons, which we will drop, however), i.e., $P_{1,2}^2 = \mathbb{I}$. The last relation leads to $P_{1,2}^{-1} = P_{1,2}$ and consequently $P_{1,2}^\dagger = P_{1,2}$, i.e., $P_{1,2}$ is also Hermitian. Its eigenvalues are thus real and their square is 1, leaving only possible eigenvalues ± 1.

As the particles's swapping must not have any impact on two-particle observables, $P_{1,2}$ commutes with all of them, in particular with the Hamiltonian. There is consequently a common eigensystem, i.e., a valid two-particle state can be written as an eigenstate of $P_{1,2}$ and the corresponding eigenvalue ± 1 remains constant under time evolution.

The phase ± 1 of a particle exchange can further not depend on the states the two particles are in. This can be shown via a contradiction:

1. Suppose the exchange is even for states a and b

$$P_{1,2}|\psi_{a,b}\rangle = +|\psi_{a,b}\rangle \tag{8.1}$$

and odd for states c and d

$$|\psi_{c,d}\rangle = \frac{1}{\sqrt{2}}(|c\rangle \otimes |d\rangle - |d\rangle \otimes |c\rangle) \quad \Rightarrow \quad P_{1,2}|\psi_{c,d}\rangle = -|\psi_{c,d}\rangle \tag{8.2}$$

2. As quantum mechanics is a linear theory, the linear combination $|\psi\rangle = |\psi_{a,b}\rangle + |\psi_{c,d}\rangle$ should also be a valid two-particle state.
3. However,

$$P_{1,2}\left(|\psi_{a,b}\rangle + |\psi_{c,d}\rangle\right) = \left(|\psi_{a,b}\rangle - |\psi_{c,d}\rangle\right) \ , \tag{8.3}$$

is not an eigenstate of $P_{1,2}$ and thus in fact not a valid two-particle state.
4. Getting invalid two-particle states with undefined symmetry can only be avoided if the phase does not depend on the states of the particles.

Moreover, it turns out that if there are more than two indistinguishable particles, exchanging any two of them always gives the same sign. In fact, relativistic quantum field theory can show that wave functions for all particles with integer spin are symmetric with respect to particle exchange while those with half-integer spin pick up a minus sign. The first kind is called "bosons" and the second "fermions".

8.1.3 Two non-relativistic electrons

Let us revisit the case of two non-relativistic (i.e., there is no coupling between the spin degree of freedom and the position-space wave function) spin$-1/2$ particles, e.g., electrons. In Sec. (3.4.1), we had looked at two ways to write the basis for two spins 1/2: The tensor-product basis (specifying the quantum state of each spin) and the singlet/triplet basis (specifying the quantum numbers of the total spin). For distinguishable spins (e.g. an electron and a neutron, or two electrons with clearly distinct position quantum numbers) both bases describe valid states and both have their uses. For *in*distinguishable spins, on the other hand, a legal state has to be an antisymmetric combination.

This observation can be reformulated and completed to discuss identical particles in different position-space states: The combined wave function of two identical spin-1/2 particles (e.g. two electrons) occupying has to be antisymmetric, because they are fermions:

$$\psi(\vec{r}_1, \vec{r}_2, \sigma_{a,1}, \sigma_{b,2}) = \psi_a(\vec{r}_1)\psi_b(\vec{r}_2)\sigma_{a,1}\sigma_{b,2} - \psi_a(\vec{r}_2)\psi_b(\vec{r}_1)\sigma_{a,2}\sigma_{b,1} \ , \tag{8.4}$$

where numbers $1, 2$ denote the particles and letters a, b the two states $|i\rangle = |\psi_i\rangle \otimes |\sigma_i\rangle$. Depending on the spin quantum numbers, there are four such states

$$\begin{aligned}
&\frac{1}{\sqrt{2}}\left(\psi_a(\vec{r}_1)\psi_b(\vec{r}_2)\uparrow_1\uparrow_2 - \psi_a(\vec{r}_2)\psi_b(\vec{r}_1)\uparrow_1\uparrow_2\right) = \frac{1}{\sqrt{2}}\left(\psi_a(\vec{r}_1)\psi_b(\vec{r}_2) - \psi_a(\vec{r}_2)\psi_b(\vec{r}_1)\right)\uparrow_1\uparrow_2 \\
&\frac{1}{\sqrt{2}}\left(\psi_a(\vec{r}_1)\psi_b(\vec{r}_2)\downarrow_1\downarrow_2 - \psi_a(\vec{r}_2)\psi_b(\vec{r}_1)\downarrow_1\downarrow_2\right) = \frac{1}{\sqrt{2}}\left(\psi_a(\vec{r}_1)\psi_b(\vec{r}_2) - \psi_a(\vec{r}_2)\psi_b(\vec{r}_1)\right)\downarrow_1\downarrow_2 \\
&\frac{1}{\sqrt{2}}\left(\psi_a(\vec{r}_1)\psi_b(\vec{r}_2)\uparrow_1\downarrow_2 - \psi_a(\vec{r}_2)\psi_b(\vec{r}_1)\downarrow_1\uparrow_2\right) = |\ A\ \rangle
\end{aligned}$$

$$\frac{1}{\sqrt{2}}\left(\psi_a(\vec{r}_1)\psi_b(\vec{r}_2)\downarrow_1\uparrow_2 - \psi_a(\vec{r}_2)\psi_b(\vec{r}_1)\uparrow_1\downarrow_2\right) = \mid B \rangle \ .$$

The first two clearly are triplet states with total spin 1, they have symmetric spin-part and antisymmetric position-part. As spin and position-space degrees of freedom do not couple, we expect to be able to find a basis, where all states are a tensor products of position-space wave function and spin state. Indeed, the sum resp. difference of last two wave functions also have such a form

$$\begin{aligned}\frac{1}{\sqrt{2}}\left(\mid A \rangle + \mid B \rangle\right) &= \frac{1}{2}\left(\psi_a(\vec{r}_1)\psi_b(\vec{r}_2)\uparrow_1\downarrow_2 - \psi_a(\vec{r}_2)\psi_b(\vec{r}_1)\downarrow_1\uparrow_2 + \psi_a(\vec{r}_1)\psi_b(\vec{r}_2)\downarrow_1\uparrow_2 - \psi_a(\vec{r}_2)\psi_b(\vec{r}_1)\uparrow_1\downarrow_2\right) = \\ &= \frac{1}{\sqrt{2}}\Big(\psi_a(\vec{r}_1)\psi_b(\vec{r}_2) - \psi_a(\vec{r}_2)\psi_b(\vec{r}_1)\Big)\frac{1}{\sqrt{2}}\Big(\uparrow_1\downarrow_2 + \downarrow_1\uparrow_2\Big) \qquad (8.5)\\ \frac{1}{\sqrt{2}}\left(\mid A \rangle - \mid B \rangle\right) &= \frac{1}{2}\left(\psi_a(\vec{r}_1)\psi_b(\vec{r}_2)\uparrow_1\downarrow_2 - \psi_a(\vec{r}_2)\psi_b(\vec{r}_1)\downarrow_1\uparrow_2 - \psi_a(\vec{r}_1)\psi_b(\vec{r}_2)\downarrow_1\uparrow_2 + \psi_a(\vec{r}_2)\psi_b(\vec{r}_1)\uparrow_1\downarrow_2\right) = \\ &= \frac{1}{\sqrt{2}}\Big(\psi_a(\vec{r}_1)\psi_b(\vec{r}_2) + \psi_a(\vec{r}_2)\psi_b(\vec{r}_1)\Big)\frac{1}{\sqrt{2}}\Big(\uparrow_1\downarrow_2 - \downarrow_1\uparrow_2\Big) \qquad (8.6)\end{aligned}$$

The first of these combined wave functions is again a triplet (i.e., symmetric in spin space) with an antisymmetric position-space wave function. The second, on the other hand, is the singlet, which is anti-symmetric in spin space and consequently requires a symmetric position-space wave function to yield an antisymmetric total wave function.

The necessity to combine an antisymmetric (i.e. singlet) spin part with a symmetric position-space part or a symmetric (i.e. triplet) spin with anti-symmetric position-space wave function effectively couples position space and spin, even for non-relativistic electrons. Two important aspects are here

1. An anti-symmetric position-space wave function $\psi_a(\vec{r}_1)\psi_b(\vec{r}_2) - \psi_a(\vec{r}_2)\psi_b(\vec{r}_1)$ vanishes for $\vec{r}_1 = \vec{r}_2$. Electrons forming a triplet consequently do not feel as strong a mutual Coulomb repulsion, because their wave functions evade each other in position space. This promotes ferromagnetic alignment and is the origin on Hund's rule coupling.
2. Two electrons can only be in the same one-particle space if they are in a singlet state. If the energy splitting between the one-particle states is large, this gives a substantial advantage to singlet states. The ground state of Helium is given by two electrons in the lowest hydrogen-like orbital, for example. In a solid, this can promote antiferromagnetism.

Sakurai has a somewhat similar discussion

For illustration, take the example of two electrons in the Helium atom. Its ground state is given by the two electrons sitting in the lowest hydrogen-like orbital, because the energy difference to the next higher state would cost more than Coulomb energy could gain from a reduced overlap. Excited states can be built by putting one electron in the lowest orbital and the second in a higher one. As Helium is a very light element, explicit spin-orbit coupling is small and does not explicitly enter the Hamiltonian. Nevertheless, as mentioned above, the different position-space symmetry of triplet and singlet wave functions gives them a different energy.

The Hamiltonian for the two electrons in Helium is given by

$$H = \underbrace{\sum_{\alpha=1,2} \frac{\vec{p}_\alpha^2}{2m} - \sum_{\alpha=1,2} \frac{2e^2}{|r_\alpha|}}_{=H_0} + \underbrace{\frac{e^2}{|\vec{r}_1 - \vec{r}_2|}}_{=H_1} , \tag{8.7}$$

i.e. the sum of the kinetic energy (first term), the potential energy due to the nucleus (second term) and the electron-electron Coulomb interaction. In principle, there might also be magnetic interactions between the spins of the nucleus and the electrons of between the electron spins, however, these are orders of magnitude smaller than the effects discussed here and can consequently be neglected.

This Hamiltonian cannot be solved exactly, and we will here use perturbation theory, where the unperturbed H_0 can be solved exactly,[1] as it is very close to two copies of the hydrogen atom, and the perturbation H_1 is the electron-electron interaction. To solve H_0, note that it is a "one-particle" operator, because it can be written as a sum of operators only acing on the coordinates of one particle, i.e.,

$$H_0 = \sum_\alpha H^\alpha = \sum_\alpha \left(\frac{\vec{p}_\alpha^2}{2m} - \frac{2e^2}{|r_\alpha|} \right) . \tag{8.8}$$

The only difference between H^α and the hydrogen Hamiltonian is the factor of 2 in the numerator of the second term (due to the double nuclear charge $Z = 2$ instead of $Z = 1$), it can be absorbed into a renormalization of the typical length $a_0 \to a_0/Z$. The one-electron ground state of H_α is denoted by $|\psi_I\rangle$, the two-electron ground state of H_0 is found by putting two electrons into this state $|\psi_I\rangle^1 \otimes |\psi_I\rangle^2$, they must have opposite spin, see second comment above.

The ground state of the full Hamiltonian can be treated in perturbation theory [2], and is still given by the singlet state. This is valid as long as Coulomb repulsion H_1 is "small" compared to H_0. Otherwise, it might lower the total energy to put one electron into a higher-energy state, thereby reducing the wave-function overlap and Coulomb repulsion. Such a ground state is qualitatively different and cannot be reached by perturbation theory. (However, in He, H_1 is small compared to H_0.)

We are here going to focus on excited states. Excited states of H_0 can be obtained by putting one (or two, but we are not going to discuss these) electrons into higher-energy states. Depending on the total spin of the two electrons, the position-space part of the singlet/triplet wave function of such an excited state is given by Eqs. (8.6) and (8.5):

$$|\psi\rangle^{S/T} = \frac{1}{\sqrt{2}} \left(|\psi_I\rangle^1 \otimes |\psi_{II}\rangle^2 \pm |\psi_{II}\rangle^1 \otimes |\psi_I\rangle^2 \right) , \tag{8.9}$$

where I and II denote states and 1, 2 particles. As $|\psi_{I/II}\rangle$ are both eigenstates of H_0, the total unperturbed energy is

$${}^{S/T}\langle\psi|H_0|\psi\rangle^{S/T} = \frac{1}{2}\Big({}^2\langle\psi_{II}| \, {}^1\langle\psi_I| \pm {}^2\langle\psi_I| \, {}^1\langle\psi_{II}| \Big)(H_0^1 + H_0^2)\Big(|\psi_I\rangle^1 |\psi_{II}\rangle^2 \pm |\psi_{II}\rangle^1 |\psi_I\rangle^2 \Big) =$$

[1] Which we will, however, not do.
[2] A variational treatment is another option.

$$
\begin{aligned}
&= \frac{1}{2}\Big(\underbrace{{}^2\langle\psi_{II}|\psi_{II}\rangle^2}_{=1}\,{}^1\underbrace{\langle\psi_I|H_0^1|\psi_I\rangle^1}_{=E_{0,I}} + {}^2\langle\psi_{II}|H_0^2|\psi_{II}\rangle^2\,{}^1\langle\psi_I|\psi_I\rangle^1 \\
&\quad + {}^2\langle\psi_I|\psi_I\rangle^2\,{}^1\langle\psi_{II}|H_0^1|\psi_{II}\rangle^1 + {}^2\langle\psi_I|H_0^2|\psi_I\rangle^2\,{}^1\langle\psi_{II}|\psi_{II}\rangle^1 \\
&\quad \pm \underbrace{{}^2\langle\psi_I|\psi_{II}\rangle^2}_{=0}\,{}^1\underbrace{\langle\psi_{II}|H_0^1|\psi_I\rangle^1}_{=E_{0,I}\cdot 0} \pm {}^2\langle\psi_I|H_0^2|\psi_{II}\rangle^2\,{}^1\langle\psi_{II}|\psi_I\rangle^1 \\
&\quad \pm {}^2\langle\psi_{II}|\psi_I\rangle^2\,{}^1\langle\psi_I|H_0^1|\psi_{II}\rangle^1 + {}^2\langle\psi_{II}|H_0^2|\psi_I\rangle^2\,{}^1\langle\psi_I|\psi_{II}\rangle^1\Big) = \\
&= E_{0,I} + E_{0,II}\ , \qquad (8.10)
\end{aligned}
$$

as expected for a state where the two levels I and II are occupied. ${}^S\langle\psi|H_0|\psi\rangle^T = 0$, because H_0 is symmetric under exchange of particles and consequently preserves the wave function's symmetry.

In first order perturbation theory, we now calculate the energy correction due to the interaction, see (5.17), which gives the terms

$$
\begin{aligned}
{}^{S/T}\langle\psi|H_1|\psi\rangle^{S/T} &= \frac{1}{2}\Big({}^2\langle\psi_{II}|\,{}^1\langle\psi_I| \pm {}^2\langle\psi_I|\,{}^1\langle\psi_{II}|\Big)H_1\Big(|\psi_I\rangle^1|\psi_{II}\rangle^2 \pm |\psi_{II}\rangle^1|\psi_I\rangle^2\Big) = \\
&= \frac{1}{2}\Big({}^2\langle\psi_{II}|\,{}^1\langle\psi_I|H_1|\psi_I\rangle^1|\psi_{II}\rangle^2 + {}^2\langle\psi_I|\,{}^1\langle\psi_{II}|H_1|\psi_{II}\rangle^1|\psi_I\rangle^2 \\
&\qquad \pm {}^2\langle\psi_{II}|\,{}^1\langle\psi_I|H_1||\psi_{II}\rangle^1|\psi_I\rangle^2 \pm {}^2\langle\psi_I|\,{}^1\langle\psi_{II}|H_1|\psi_I\rangle^1|\psi_{II}\rangle^2\Big)\ .
\end{aligned}
$$

The two terms on the last line give the same, as do the two on the second-last, because H_1 is the same regardless of which particle sits in state I and which in II, leaving two terms

$$
{}^{S/T}\langle\psi|H_1|\psi\rangle^{S/T} = \Big(\underbrace{{}^2\langle\psi_{II}|\,{}^1\langle\psi_I|H_1|\psi_I\rangle^1|\psi_{II}\rangle^2}_{\text{"direct"}} \pm \underbrace{{}^2\langle\psi_I|\,{}^1\langle\psi_{II}|H_1|\psi_I\rangle^1|\psi_{II}\rangle^2}_{\text{"exchange"}}\Big)\ . \qquad (8.11)
$$

The first, called "direct", has the form one would intuitively expect for a Coulomb interaction between charge densities $e|\psi|^2$

$$
\begin{aligned}
I_{\text{direct}} &= \int d^3r_1 \int d^3r_2\ \psi_{II}^*(\vec{r}_2)\psi_I^*(\vec{r}_1)\frac{e^2}{|\vec{r}_1-\vec{r}_2|}\psi_{II}(\vec{r}_2)\psi_I(\vec{r}_1) = \\
&= \int d^3r_1 \int d^3r_2\ |\psi_{II}(\vec{r}_2)|^2|\psi_I(\vec{r}_1)|^2\frac{e^2}{|\vec{r}_1-\vec{r}_2|} > 0\ . \qquad (8.12)
\end{aligned}
$$

The second "exchange" term does not feature densities $|\psi|^2$ and is the more exotic result of exchange symmetry:

$$
I_{\text{exchange}} = \int d^3r_1 \int d^3r_2\ \psi_{II}^*(\vec{r}_2)\psi_I^*(\vec{r}_1)\frac{e^2}{|\vec{r}_1-\vec{r}_2|}\psi_{II}(\vec{r}_1)\psi_I(\vec{r}_2)\ . \qquad (8.13)
$$

While it is not obvious, $I_{\text{exchange}} > 0$ as well. As the antisymmetric position-space wave function of the triplet state goes together with the minus sign, its energy is $2I_{\text{exchange}}$ lower than that of the singlet state. This splitting is not due to any explicit spin-spin coupling, but to Coulomb repulsion together with the requirement that the wave function be totally antisymmetric.

8.2 Symmetrized Many-body States: Generalities and Bosons

In order to treat particles as indistinguishable, the total wave function should be (anti-) symmetric under exchange of any two particles for bosons (fermions). This can be done by averaging over all permutations $\mathcal{P}$ that mix around the particles in their states. A general permutation can always be written as a product of "swaps" like $P_{1,2}$ discussed above. For fermions, each swap adds a sign (-1) to the wave function and the total permutation $\mathcal{P}$ then carries a sign $(-1)^{m_\mathcal{P}}$, where $m_\mathcal{P}$ is the number of swapping operators needed. (While the decomposition of a permutation into swaps is not unique, the total sign is.)

The (anti-)symmetrized wave function $|\psi\rangle^{A/S}$ is proportional to the superposition of all the permutations:

$$|\psi\rangle^{A/S} \propto \frac{1}{\sqrt{N!}} \sum_\mathcal{P} (\pm 1)^{m_\mathcal{P}} \; \mathcal{P} \; |i_1\rangle^1 \otimes |i_2\rangle^2 \cdots \otimes |i_N\rangle^N \; . \tag{8.14}$$

$|i_1\rangle^1$ denotes here that particle 1 (superscript as well as position in the total state) is in state $|i_1\rangle$, one permutation would for example be $P_{1,2}|i_1\rangle^1 \otimes |i_2\rangle^2 \cdots \otimes |i_N\rangle^N = |i_2\rangle^1 \otimes |i_1\rangle^2 \cdots \otimes |i_N\rangle^N$. In $(\pm 1)^{m_\mathcal{P}}$, the $+$ sign applies to bosons and the $-$ sign to fermions. However, it turns out that this state is not yet correctly normalized. The normalization is only correct if all particle are in different states, it.e., if each state is at most singly occupied [3]. If we have n_i particles in one state i, then, the symmetrized wave function above reads

$$\frac{1}{\sqrt{n_i!}} \sum_\mathcal{P} \mathcal{P} \; |i\rangle^1 \otimes |i\rangle^2 \otimes \ldots |i\rangle^{n_i} = \sqrt{n_i!} \; |i\rangle^1 \otimes |i\rangle^2 \otimes \ldots |i\rangle^{n_i} \; , \tag{8.15}$$

i.e., only $\sqrt{n_i!}$ rather than $n_j!$ is taken into account by the normalization. One can verify that this remains the same if there are additional particles in additional states; a factor of $\sqrt{n_i!}$ arises for each state. The normalized wave function is then

$$|\psi\rangle^{A/S} = \frac{1}{\sqrt{N!}} \frac{1}{\sqrt{n_1! n_2! \ldots n_M!}} \sum_\mathcal{P} (\pm 1)^{m_\mathcal{P}} \; \mathcal{P} \; |i_1\rangle^1 \otimes |i_2\rangle^2 \cdots \otimes |i_N\rangle^N \; , \tag{8.16}$$

where n_k is the number of particles in state $|i_k\rangle$.

One can check that this state is orthonormal if the $|i_k\rangle$ are, i.e., ${}^S\langle\psi|\tilde{\psi}\rangle^S = 1$ if the states are equivalent in the sense that (unsymmetrized) $|\psi\rangle$ is one of the permutations of (unsymmetrized) $|\tilde{\psi}\rangle$, and ${}^S\langle\psi|\tilde{\psi}\rangle^S = 0$ otherwise [4].

Working with such (anti-)symmetrized wave functions is rather tedious for more than two particles. Moreover, it seems wasteful to carefully write down detailed information (which particles sits in with state) only to then average over it. It would be preferable to have a formalism, where we only need to specify the number of particles in each state, as this is the only physically relevant information. Wave functions like (8.16) for N particles are then written as $|n_I, n_{II}, \ldots n_M\rangle$ with $\sum_k n_k = N$. In fact, we can leave off the restriction $\sum_k n_k = N$ and instead include arbitrary N into a so-called "Fock" space that is the direct sum of the Hilbert spaces for 0, 1, ... identical particles. The symmetrized wave function (8.16) can be *written* using the particle numbers, but what remains to be shown is that

[3] The normalization is thus fine for fermions.

[4] For fermions, ${}^A\langle\psi|\tilde{\psi}\rangle^A = -1$ is also possible and counts as equivalent to 1

we can build a working formalism using only particle numbers, i.e., without having to use (8.16).

A first thing to check are basic properties of the states, e.g., that these states are orthonormalized. The requirement that (unsymmetrized) $|\psi\rangle$ be a permutation of (unsymmetrized) $|\tilde{\psi}\rangle$ translates to having the same particle numbers $n_I = \tilde{n}_i \ldots$, i.e.,

$$\langle n_I, n_{II}, \ldots n_M | \tilde{n}_I, \tilde{n}_{II}, \ldots \tilde{n}_M \rangle = \delta_{n_I,\tilde{n}_I} \delta_{n_{II},\tilde{n}_{II}} \ldots \delta_{n_M,\tilde{n}_M} \ . \tag{8.17}$$

As any fully (anti-)symmetric state of the Fock space can be expressed in terms of states (8.16), they also form a basis

$$\mathbb{I} = \sum_{n_I, n_{II}, \ldots, n_M} |n_I, n_{II}, \ldots n_M\rangle\langle n_I, n_{II}, \ldots n_M| \tag{8.18}$$

W have already seen states of such a structure, namely the eigenstates of the Hamiltonian describing coupled harmonic oscillators, see Eq. (7.12). Building on this, we postulate an analogous formalism here. Just as $a_k^\dagger$ and a_k, which mediate between eigenstates with N_k energy quanta in mode k and eigenstates with $N_k \pm 1$ quanta, were interpreted to create/annihilate phonons, $b_i^\dagger$ and b_i are now introduced to create and annihilate a boson on state $|i\rangle$. Their action on symmetrized wave function is defined as:

$$b_i^\dagger |n_I, n_{II}, \ldots, n_i, \ldots, n_M\rangle = \sqrt{n_i + 1}\ |n_I, n_{II}, \ldots, n_i + 1, \ldots, n_M\rangle \quad \text{and} \tag{8.19}$$

$$b_i |n_I, n_{II}, \ldots, n_i, \ldots, n_M\rangle = \sqrt{n_i}\ |n_I, n_{II}, \ldots, n_i - 1, \ldots, n_M\rangle \ . \tag{8.20}$$

Note that b_i gives 0 when applied to a state with $n_i = 0$.

Even though these new boson operators were not derived from position and momentum operators, we still find the same commutation properties:

$$\begin{aligned} b_j^\dagger b_i^\dagger |n_I, n_{II}, \ldots, n_i, \ldots, n_j, \ldots, n_M\rangle &= b_j^\dagger \sqrt{n_i + 1} |n_I, n_{II}, \ldots, n_i + 1, \ldots, n_j, \ldots, n_M\rangle = \\ &= \sqrt{n_j + 1}\sqrt{n_i + 1} |n_I, n_{II}, \ldots, n_i + 1, \ldots, n_j + 1, \ldots, n_M\rangle = \\ = b_i^\dagger b_j^\dagger |n_I, n_{II}, \ldots, n_i, \ldots, n_j, \ldots, n_M\rangle & \end{aligned} \tag{8.21}$$

and analogously $b_i b_j = b_j b_i$. For $i \neq j$, b_i and $b_j^\dagger$ also commute, but for $i = j$,

$$\begin{aligned} b_i b_i^\dagger |n_I, n_{II}, \ldots, n_i, \ldots, n_M\rangle &= b_i \sqrt{n_i + 1} |n_I, n_{II}, \ldots, n_i + 1, \ldots, n_M\rangle = \\ &= (n_i + 1) |n_I, n_{II}, \ldots, n_i, \ldots, n_M\rangle \quad \text{while} \\ b_i^\dagger b_i |n_I, n_{II}, \ldots, n_i, \ldots, n_M\rangle &= b_i^\dagger \sqrt{n_i} |n_I, n_{II}, \ldots, n_i - 1, \ldots, n_M\rangle = \\ &= n_i\ |n_I, n_{II}, \ldots, n_i, \ldots, n_M\rangle \ . \end{aligned} \tag{8.22}$$

We note here that the many-boson state is an eigenstate of $b_i^\dagger b_i$ (as well as of $b_i b_i^\dagger$), with the particle number n_i ($n_i + 1$) as the eigenvalue:

$$b_i^\dagger b_i |n_I, \ldots, n_i, \ldots\rangle = \hat{n}_i\ |n_I, \ldots, n_i, \ldots\rangle = n_i\ |n_I, \ldots, n_i, \ldots\rangle \ . \tag{8.23}$$

The operator $\hat{n}_i$ is the number operator for state $|i\rangle$. The commutation relations can be summarized to

$$[b_i, b_j] = [b_i^\dagger, b_j^\dagger] = 0, \quad [b_i, b_j^\dagger] = \delta_{i,j} \ . \tag{8.24}$$

The basis states of the Fock-space can then be obtained by repeated application of $b^\dagger$, e.g., the one-particle states $|i\rangle = b_i^\dagger|0\rangle$ and two-particle states are $b_i^\dagger b_j^\dagger|0\rangle$ or $\frac{1}{\sqrt{2}}(b_i^\dagger)^2|0\rangle$. General states with $N = \sum_n n_k$ particles look just like (7.12), i.e.

$$|\psi\rangle^S = |n_I, n_{II}, \ldots, n_M\rangle = \tag{8.25}$$
$$= \frac{1}{\sqrt{n_I!}}\left(b_I^\dagger\right)^{n_I}\frac{1}{\sqrt{n_{II}!}}\left(b_{II}^\dagger\right)^{n_{II}}\cdots\frac{1}{\sqrt{n_M!}}\left(b_M^\dagger\right)^{n_M}|0\rangle = \prod_k \frac{1}{\sqrt{n_k!}}\left(b_k^\dagger\right)^{n_k}|0\rangle\ .$$

8.2.1 Operators in Second Quantization

Having a formalism to write symmetrized states would not be very useful if we can not translate operators into this language. An important special type of operators in many-particle systems are "one-particle" operators. They act on each particle separately, i.e., in "first quantization" language, they are written as a sum of operators acting only on the wave function of a single particle, see for example Eq. (8.8). As the particles are indistinguishable, all terms in the sum have to look the same, i.e., for operator T^α acting on particle α, the many-body version is:

$$T^\alpha = \sum_{i,j} t_{i,j}|i\rangle^{\alpha\,\alpha}\langle j| \quad \Rightarrow \quad \hat{T} = \sum_\alpha T^\alpha = \sum_{i,j} t_{i,j}\sum_\alpha |i\rangle^{\alpha\,\alpha}\langle j|\ . \tag{8.26}$$

An example would be the kinetic energy of a group of identical particles, which is given by the the sum over the kinetic energies of the individual particles $\sum_\alpha \frac{\hat{p}_\alpha^2}{2m}$. Another example is the energy due to an external potential acting on the particles $\sum_\alpha V(\hat{\vec{r}}_\alpha)$.

Now, we apply one term to the symmetrized wave function; we are interested in its action on (8.25), but as the operator is written in terms of one-particle wave functions, we apply it to the "first quantization" version of the many-body state, to (8.16):

$$t_{i,j}\sum_\alpha |i\rangle^{\alpha\,\alpha}\langle j|\psi\rangle^S = t_{i,j}\sum_\alpha |i\rangle^{\alpha\,\alpha}\langle j|n_I, n_{II}, \ldots, n_i, \ldots, n_j, \ldots, n_M\rangle = \tag{8.27}$$
$$= t_{i,j}\frac{1}{\sqrt{N!}}\frac{1}{\sqrt{n_1!n_2!\ldots n_M!}}\sum_{\mathcal{P}}\mathcal{P}\sum_\alpha |i\rangle^{\alpha\,\alpha}\langle j|\ |i_1\rangle^1\otimes|i_2\rangle^2\cdots\otimes|i_N\rangle^N\ ,$$

where $|i\rangle^{\alpha\,\alpha}\langle j|$ gives 0 if particle α is not in state $|j\rangle$ and puts it into state $|i\rangle$ if it is. This means replacing $|i_\alpha\rangle^\alpha$ by $|i_\alpha = i\rangle\delta_{i_\alpha,j}$. For each perturbation $\mathcal{P}$, the n_j particles in state $|j\rangle$ must be at *some* positions β_j, the sum running over α will consequently n_j times pick up a non-zero contribution. Every time this happens, one particle is moved from $|j\rangle$ to $|i\rangle$, i.e., all contributions to the new state have state $|i\rangle$ occupied by one more particle than before and $|j\rangle$ by one fewer. The state then becomes

$$t_{i,j}n_j\frac{1}{\sqrt{N!}}\frac{1}{\sqrt{n_1!n_2!\ldots n_M!}}\sum_{\mathcal{P}}\mathcal{P}\underbrace{|i_1\rangle^1\otimes|i_2\rangle^2\cdots\otimes|i\rangle^{\beta_j}\cdots\otimes|i_N\rangle^N}_{\substack{\text{one more in}|i\rangle\\ \text{one fewer in}|j\rangle}}\ . \tag{8.28}$$

Except for the normalization, this is a symmetrized wave function with $n_i \to n_i + 1$ and

$n_j \to n_j - 1$, fixing the normalization gives

$$t_{i,j} n_j \underbrace{\frac{\sqrt{(n_i+1)!}}{\sqrt{n_i!}}}_{\sqrt{n_i+1}} \underbrace{\frac{\sqrt{(n_j-1)!}}{\sqrt{n_j!}}}_{\sqrt{n_j}^{-1}} \times$$
$$\times \frac{1}{\sqrt{N!}} \frac{1}{\sqrt{n_1! n_2! \ldots (n_i+1)! \ldots (n_j-1)! \ldots n_M!}} \sum_{\mathcal{P}} \mathcal{P} |i_1\rangle^1 \otimes |i_2\rangle^2 \cdots \otimes |i_N\rangle^N$$
$$= t_{i,j} \sqrt{n_j} \sqrt{n_i+1} |n_I, n_{II}, \ldots, n_i+1, \ldots, n_j-1, \ldots, n_M\rangle \; . \tag{8.29}$$

Comparing to definitions (8.20) and (8.19) shows that this state is equivalent to

$$t_{i,j} b_i^\dagger b_j |n_I, n_{II}, \ldots, n_i, \ldots, n_j, \ldots, n_M\rangle \; . \tag{8.30}$$

As the equivalence holds for all symmetrized states and as they form a basis of the Fock space of allowed many-boson states, the operators must be the same and one-particle operator (8.26) in second quantization is

$$\hat{T} = \sum_{i,j} t_{i,j} \; b_i^\dagger b_j \; . \tag{8.31}$$

The matrix element $t_{i,j}$ is the same as in the operator T^1 for one single particle, i.e., $t_{i,j} = \langle i | T^1 | j \rangle$, which is in position space

$$t_{i,j} = \int d^3r \; \psi_i^*(\vec{r}) \; \hat{T}^1 \; \psi_j(\vec{r}) \; . \tag{8.32}$$

One can understand the action of the (many-body) "one-particle" operator as taking a particle from state j and putting it into state i with an amplitude $\langle i | T^1 | j \rangle$ that is the "overlap" between the states $|i\rangle$ and $T^1 |j\rangle$.

One-particle operators discussed so far treat the many particles as if they were independent - the only restriction is the requirement of a symmetric wave function, which is already enforced by the formalism of "second quantization". In "first quantization", true interactions between particles are expressed by operators acting on the coordinates of more than one particle. An example would be Coulomb repulsion between two particles, $\propto 1/|\hat{\vec{r}}_1 - \hat{\vec{r}}_2|$, where the position operators $\hat{\vec{r}}_1$ and $\hat{\vec{r}}_2$ of both enter. The most common form are two-particle interactions, which can be written as

$$F = \frac{1}{2} \sum_{\alpha \neq \beta} V_{i,i';j,j'} |i\rangle^\alpha \otimes |i'\rangle^\beta \, {}^\beta\langle j'| \otimes {}^\alpha\langle j| \; , \tag{8.33}$$

where the factor $1/2$ ensures that every pair of particles enters only once [5]. The two-particle matrix element $V_{i,i',j',j}$ can be calculated with unsymmetrized wave functions, e.g. in position space with the integral

$$V_{i,i';j,j'} = {}^2\langle i'| \otimes {}^1\langle i| \; \hat{V}^{1,2} |j\rangle^1 \otimes |j'\rangle^2 = \int d^3r_1 \int d^3r_2 \; \psi_i^*(\vec{r}_1) \psi_{i'}^*(\vec{r}_2) \; \hat{V}^{1,2} \; \psi_j(\vec{r}_1) \psi_{j'}(\vec{r}_2) \; . \tag{8.34}$$

[5] Alternatively, one could write the sum to go over $\alpha < \beta$ only.

We have seen such matrix elements for Coulomb interaction in Eqs. (8.12) and (8.13). The operator $\hat{V}^{1,2} = \hat{V}(\hat{\vec{r}}_1, \hat{\vec{r}}_1, \dots)$ might in addition to the positions of the two particles also depend on their momenta, spin ...

Similar arithmetic as above leads to the formulation in second quantization

$$F = \frac{1}{2} \sum_{i,i';j,j'} V_{i,i';j,j'}\ b_i^\dagger b_{i'}^\dagger b_{j'} b_j = \frac{1}{2} \sum_{i,i';j>j'} (V_{i,i';j,j'} + V_{i',i;j',j}) b_i^\dagger b_{i'}^\dagger b_{j'} b_j = \sum_{i,i';j>j'} V_{i,i';j,j'}\ b_i^\dagger b_{i'}^\dagger b_{j'} b_j \ , \tag{8.35}$$

where $V_{i,i';j,j'} = V_{i',i;j',j}$ (see (8.34) and $b_i^\dagger b_{i'}^\dagger b_{j'} b_j = b_{i'}^\dagger b_i^\dagger b_j b_{j'}$ (see (8.24; also valid for fermions, see (8.49)) were used. A two particle operator thus takes two particles from their states and puts them into other states, again with an amplitude related to an overlap.

8.2.2 Alternative derivation of one-particle operators: basis change

following Sakurai

There is an illustrative alternative way to obtain one-particle operators. Let us assume that operator T^1 is diagonal in the basis used to define the creation/annihilation operators, i.e., the states $|i\rangle$, $|j\rangle$,... are eigenstates of $\hat{T}^1 = \sum_i t_i |i\rangle\langle i|$. The many-body operator $\hat{T} = \sum_\alpha \hat{T}^\alpha$ then has a very simple form: As T^1 does not not change the states of the particles, $\sum_\alpha \hat{T}^\alpha$ conserves all occupation numbers n_i. As seen in (8.23), the many-boson states are simultaneous eigenstates of all particle-number operators $\hat{n}_i$ (with eigenvalue n_i) and consequently also of $\hat{T}$. Applying $\hat{T}$ to the many-boson state just adds t_i for every particle in state i:

$$\hat{T} = \sum_i t_i n_i \tag{8.36}$$

An example for such an operator is the Hamiltonian of the harmonic oscillator (1.51) with eigenvalue (1.59) and its variants for phonons, (7.10) with eigenvalue (7.13), and photons, (7.52) with eigenvalue (7.53).

In general, T^1 can of course not be expected to be diagonal, but we might change the basis to make it so. Here, $|b_i\rangle$ and $|a_i\rangle$ will be used to denote states of the two bases and for a single particle, a basis change is done via

$$|a_i\rangle = \sum_j u_{j,i} |b_j\rangle = \sum_j |b_j\rangle\langle b_j|a_i\rangle \ . \tag{8.37}$$

Just as $b_i^\dagger|0\rangle = |b_i\rangle$, we introduce operators $a_i^\dagger|0\rangle = |a_i\rangle$ for bosons in the new basis. This leads to the definition of a basis change [6] for creation and annihilation operators

$$a_i^\dagger = \sum_j b_j^\dagger \langle b_j|a_i\rangle \quad \text{and} \quad a_i = \left(a_i^\dagger\right)^\dagger = \sum_j b_j \langle a_i|b_j\rangle \tag{8.38}$$

If $\hat{T}^1$ is diagonal in the $|a\rangle$ basis, we can make use of the simple form for the many-body operator given above and transform it back into the $|b\rangle$ basis. Making use of the fact that

[6] It is instructive to check that such a basis transformation preserves the all-important commutation relation!

n_i^a (the particle number when using the $|a\rangle$-basis) is the eigenvalue of $a_i^\dagger a_1$, see (8.22), we get

$$\hat{T} = \sum_i t_i n_i^a = \sum_i t_i a_i^\dagger a_i = \sum_i t_i \sum_{j,k} b_j^\dagger b_k \langle b_j|a_i\rangle\langle a_i|b_k\rangle = \sum_{j,k} b_j^\dagger b_k \langle b_j| \underbrace{\sum_i t_i |a_i\rangle\langle a_i|}_{=\hat{T}^1} |b_k\rangle =$$

$$= \sum_{j,k} t_{j,k}\ b_j^\dagger b_k\ . \tag{8.39}$$

We have actually already used basis transformation applied to creation/annihilation operators when going from $a_{k,s}$ and $a_{k,s}^\dagger$ in momentum space to the quantized vector potential in position space.

8.3 Anti-Symmetrized Many-body States for Fermions

For fermions, the minus sign in (8.16) applies and the contributions to the antisymmetric wave functions have positive and negative sign. Formally, the antisymmetric wave function can be written as a determinant

$$|\psi\rangle^A = \frac{1}{\sqrt{N!}} \begin{vmatrix} |i_1\rangle^1 & |i_1\rangle^2 & \dots & |i_1\rangle^N \\ |i_2\rangle^1 & |i_2\rangle^2 & \dots & |i_2\rangle^N \\ \vdots & & & \\ |i_N\rangle^1 & |i_N\rangle^2 & \dots & |i_N\rangle^N \end{vmatrix}, \tag{8.40}$$

where each row of the corresponds to a one-particle state and each column to a particle. The sign $(-1)^{m_\mathcal{P}}$ in the many-fermion wave function is the same that one finds when evaluating this so-called "Slater determinant". A few observations can be gained from the properties of determinants:

1. A determinant changes sign when two columns are switched – just like the fermion wave function upon switching two particles.
2. The determinant vanishes when two rows are the same, as the Pauli principle forbids a state to be occupied by more than one fermion.
3. It also changes sign when two rows – i.e. two states – are switched, which we will use later to define anti-commutation properties of fermion creation operators.
4. A determinant does not depend on the basis, i.e., the many-body fermion state remains a valid state if the underlying basis $|i_j\rangle$ is transformed into another one.

When extending the creator/annihilator formalism to fermions with their antisymmetric wave functions, we can recycle almost all of the previous section, but have to make sure to introduce the minus-sign of the antisymmetry correctly. Up to the introduction of the Fock space and the occupation-number notation, about until Eq. (8.18), we can apply everything to fermions as well. One technical difference has to do with the orthonormality of (8.16): For bosons, two symmetrized states are "the same" if their occupation numbers are the same, because one of the states must be a permutation of the other. For fermions, two such states might differ by a minus-sign. To remove this ambiguity and to fix the sign of a many-fermion wave function, one chooses a "canonical" order of the *states* (not

the particles!). This corresponds to writing the rows of any Slater determinant always in this order; additionally, one fixes the columns of the Slater determinant to have ascending particle index. This convention ensures that two Slater determinants either have the same sign and a scalar product on 1 or are orthogonal to each other. The occupation-number states then also form an orthonormal basis of the fermion Fock space, just like for bosons.

We would like to introduce creation operators $c_i^\dagger$ for fermions as well, which should increase the number of fermions in state $|i\rangle$ by one, i.e., $c_i^\dagger|0\rangle = |i\rangle$ and generally

$$c_i^\dagger|n_I, n_{II}, \ldots, n_i, \ldots, n_M\rangle = \begin{cases} c^m|n_I, n_{II}, \ldots, n_i + 1, \ldots, n_M\rangle & \text{if } n_i = 0 \\ 0 & \text{if } n_i = 1 \end{cases}, \tag{8.41}$$

where we do not know the prefactor c^m yet. Moreover, creation and annihilation operators need to somehow create an *anti*symmetrized wave function rather than the symmetric boson one. How this is done can be seen by making sure that occupation-number states keep a consistent sign.

Many-fermion states can be defined like many-boson states (8.25), but the order of the states needs to be fixed, as we have seen for the Slater determinant:

$$|\psi\rangle^A = |n_I, n_{II}, \ldots, n_M\rangle = \left(c_I^\dagger\right)^{n_I}\left(c_{II}^\dagger\right)^{n_{II}}\ldots|0\rangle = \prod_k \left(c_k^\dagger\right)^{n_k}|0\rangle\,. \tag{8.42}$$

For simplicity, consider only two available one-particle states, I and II. The "canonical" two-fermion state is then $|n_I = 1, n_{II} = 1\rangle = c_I^\dagger c_{II}^\dagger|0\rangle$. From the Slater determinant, which changes sign upon swapping two rows, one infers that the wave-function's sign should change if the two *states* are swapped, i.e., $c_{II}^\dagger c_I^\dagger|0\rangle = -c_I^\dagger c_{II}^\dagger|0\rangle$. As the Slater determinant suggests, this remains true for more than two states (and fermions) and is thus an operator identity:

$$c_i^\dagger c_j^\dagger = -c_j^\dagger c_i^\dagger \quad \text{resp.} \quad c_i^\dagger c_j^\dagger + c_j^\dagger c_i^\dagger = \{c_i^\dagger, c_j^\dagger\} = [c_i^\dagger, c_j^\dagger]_+ = 0 \tag{8.43}$$

Where the last expressions are two ways to denote the so-called "anti-commutator". As $c_i^\dagger c_i^\dagger + c_i^\dagger c_i^\dagger = 0$, the anti-commutation properties also enforce the Pauli principle – a sign that this scheme of making the wave function antisymmetric works.

The sign in (8.41) can now be fixed to $c^m = (-1)^{\sum_{j<i} n_j}$ and

$$\begin{aligned} c_i^\dagger|n_I, n_{II}, \ldots, n_i, \ldots, n_M\rangle &= \begin{cases} (-1)^{\sum_{j<i} n_j}|n_I, n_{II}, \ldots, n_i + 1, \ldots, n_M\rangle & \text{if } n_i = 0 \\ 0 & \text{if } n_i = 1 \end{cases} = \\ &= (1 - n_i)(-1)^{\sum_{j<i} n_j}|n_I, n_{II}, \ldots, n_i + 1, \ldots, n_M\rangle\,. \end{aligned} \tag{8.44}$$

The action of $c_i = \left(c_i^\dagger\right)^\dagger$ can then be inferred as

$$\begin{aligned} c_i|n_I, n_{II}, \ldots, n_i, \ldots, n_M\rangle &= \begin{cases} (-1)^{\sum_{j<i} n_j}|n_I, n_{II}, \ldots, n_i - 1, \ldots, n_M\rangle & \text{if } n_i = 1 \\ 0 & \text{if } n_i = 0 \end{cases} = \\ &= n_i(-1)^{\sum_{j<i} n_j}|n_I, n_{II}, \ldots, n_i - 1, \ldots, n_M\rangle \end{aligned} \tag{8.45}$$

and they anticommute like the creators. For $i < j$, we also find

$$c_i^\dagger c_j|n_I, n_{II}, \ldots, n_i, \ldots, n_j, \ldots, n_M\rangle = n_j(-1)^{\sum_{k<j} n_k} c_i^\dagger|n_I, n_{II}, \ldots, n_i, \ldots, n_j - 1, \ldots, n_M\rangle =$$

$$= n_j(-1)^{\sum_{k<j} n_k}(1-n_i)(-1)^{\sum_{k<i} n_k}|n_I, n_{II}, \ldots, n_i+1, \ldots, n_j-1, \ldots, n_M\rangle \quad \text{and} \tag{8.46}$$

$$c_j c_i^\dagger |n_I, n_{II}, \ldots, n_i, \ldots, n_j, \ldots, n_M\rangle = (1-n_i)(-1)^{\sum_{k<i} n_k} c_j |n_I, n_{II}, \ldots, n_i+1, \ldots, n_j, \ldots, n_M\rangle =$$
$$= n_j(-1)^{\sum_{k<j} n_k+1}(1-n_i)(-1)^{\sum_{k<i} n_k}|n_I, n_{II}, \ldots, n_i+1, \ldots, n_j-1, \ldots, n_M\rangle \tag{8.47}$$

The relative minus sign in $(-1)^{\sum_{k<j} n_k+1}$ arises, because the annihilator c_i has reduced the number of particles with $k<j$ in the second case. Finally, one gets

$$c_i^\dagger c_i |n_I, n_{II}, \ldots, n_i, \ldots, n_M\rangle = n_i(-1)^{\sum_{k<i} n_k} c_i^\dagger |n_I, n_{II}, \ldots, n_i-1, \ldots, n_M\rangle =$$
$$= n_i(-1)^{\sum_{k<i} n_k}(1-(n_i-1))(-1)^{\sum_{k<i} n_k}|n_I, n_{II}, \ldots, n_i, \ldots, n_M\rangle =$$
$$= n_i|n_I, n_{II}, \ldots, n_i, \ldots, n_M\rangle \quad \text{and}$$
$$c_i c_i^\dagger |n_I, n_{II}, \ldots, n_i, \ldots, n_M\rangle = (1-n_i)(-1)^{\sum_{k<i} n_k} c_i |n_I, n_{II}, \ldots, n_i+1, \ldots, n_M\rangle =$$
$$= (1-n_i)|n_I, n_{II}, \ldots, n_i, \ldots, n_M\rangle \tag{8.48}$$

where $n_i^2 = n_i$ (valid because $n_i = 0, 1$) was used. In addition to noting that the many-fermion states are eigenstates of $\hat{n}_i = c_i^\dagger c_i$ with the particle number in state $|i\rangle$ as eigenvalue, we can now summarize the anticommutation relations for fermion operators:

$$\left\{c_i^\dagger, c_j^\dagger\right\} = \left\{c_i, c_j\right\} = 0 \quad \text{and} \quad \left\{c_i^\dagger, c_j\right\} = \delta_{i,j} \tag{8.49}$$

One-particle operators in second quantization have the same form for bosons and fermions, but we should verify that there is indeed no additional minus-sign for the fermions. When applying the operator to an antisymmetrized fermion state,

$$t_{i,j}\sum_\alpha |i\rangle^\alpha {}^\alpha\langle j|\psi\rangle^A = t_{i,j}\frac{1}{\sqrt{N!}}\sum_{\mathcal{P}}(-1)^{m_{\mathcal{P}}}\mathcal{P}\sum_\alpha |i\rangle^\alpha {}^\alpha\langle j\ |i_1\rangle^1 \otimes |i_2\rangle^2 \cdots \otimes |i_N\rangle^N \tag{8.50}$$

the sum over α will for each permutation find the particle sitting in state $|j\rangle$ (if it is occupied) and replace it by a particle in state $|i\rangle$ (only allowed if $|i\rangle$ is empty). The resulting state is also a sum over all possible permutations of a fermion state, even with the correct normalization. However, state $|i\rangle$ appears wherever state $|j\rangle$ appeared in the original state, i.e., not in its canonical position. To order the state, we need to move state $|i\rangle$ to its position, i.e., anticommute $c_i^\dagger$ with all $c_k^\dagger$ with k between i and j, see (8.42). For $i<j$, this gives

$$t_{i,j}\sum_\alpha |i\rangle^\alpha {}^\alpha\langle j|\psi\rangle^A = t_{i,j}(-1)^{\sum_{i<k<j} n_k} n_j(1-n_i)|n_I, n_{II}, \ldots, n_i+1, \ldots, n_j-1, \ldots, n_M\rangle\ . \tag{8.51}$$

As can be seen from (8.46), this is the same as

$$t_{i,j} c_i^\dagger c_j |n_I, n_{II}, \ldots, n_i, \ldots, n_j, \ldots, n_M\rangle\ . \tag{8.52}$$

The equality also holds for $i>j$ and is again an operator identity, giving us as in (8.31)

$$\hat{T} = \sum_{i,j} t_{i,j}\ c_i^\dagger c_j\ . \tag{8.53}$$

Two-particle operators also remain the same

$$F = \sum_{i,i';j,j'} V_{i,i';j,j'}\ c_i^\dagger c_{i'}^\dagger c_{j'} c_j\ , \tag{8.54}$$

however, one now has to make sure to have to correct order of the operators!

8.3.1 Two non-relativistic electrons AGAIN

We will now compare the two formalisms using the example of two electrons in the Helium atom from Sec. 8.1.3, even though second quantization does here not lead to a large advantage. As discussed in Sec. 8.1.3, the ground state is obtained by putting two electrons (necessarily of different spin) into the lowest one-particle state, $|\Phi^0\rangle = c_{I,\uparrow}^\dagger c_{I,\downarrow}^\dagger |0\rangle$.

Concerning the excited states with one electron in state $|I\rangle$ and one in state $|II\rangle$, four possible excited states arise in zeroth order: $|n_{I,\uparrow}, n_{I,\downarrow}, n_{II,\uparrow}, n_{II,\downarrow}\rangle = |1,0,0,1\rangle = c_{I,\uparrow}^\dagger c_{II,\downarrow}^\dagger |0\rangle$, $|0,1,1,0\rangle = c_{I,\downarrow}^\dagger c_{II,\uparrow}^\dagger |0\rangle$, $|1,0,1,0\rangle = c_{I,\uparrow}^\dagger c_{II,\uparrow}^\dagger |0\rangle$, and $|0,1,0,1\rangle = c_{I,\downarrow}^\dagger c_{II,\downarrow}^\dagger |0\rangle$. Their zeroth-order energy is of course again

$$\sum_i E_{0,i} \sum_\sigma n_{i,\sigma} = E_{0,I} \cdot (1+0) + E_{0,II} \cdot (1+0) = E_{0,I} + E_{0,II}\ . \tag{8.55}$$

The last two of these states are clearly triplet states with spin 1. The first two are less symmetric than the Hamiltonian H, which contains no term pinning spin up to level I or II. We thus suspect that their degeneracy is "accidental", i.e., not due to any symmetry, and H_1 will split them into the missing third triplet and the singlet state.

H_1 is trivially symmetric under rotations in spin space, because it does not refer to the spin at all. S^z is consequently conserved and any terms changing it, e.g. $\propto c_{I,\uparrow}^\dagger c_{II,\uparrow}^\dagger c_{I,\uparrow} c_{II,\downarrow}$, vanish by symmetry. In fact, if we include spin space into the labels of the one-electron states (8.34) becomes here

$$\begin{aligned} {}^1\langle i,\sigma_i|\,{}^2\langle i',\sigma_{i'}|\hat{V}|j,\sigma_j\rangle^1|j',\sigma_{j'}\rangle^2 &= {}^1\langle\sigma_i|\sigma_j\rangle^1\,{}^2\langle\sigma_{i'}|\sigma_{j'}\rangle^2 \cdot {}^1\langle i|\,{}^2\langle i'|\hat{V}|j\rangle^1|j'\rangle^2 = \\ &= V_{i,i';j,j'}\delta_{\sigma_i,\sigma_j}\delta_{\sigma_{i'},\sigma_{j'}} \end{aligned} \tag{8.56}$$

In first-order perturbation theory, see Eq. (5.25), we are only interested in terms involving the four states given above, i.e., exactly one creation and one annihilation operator must refer to state I. In Eq. (8.35), this leaves only the terms with $j = I$, $j' = II$ and $i = j = I; i' = j' = II$ as well as those with $i = j' = II$ and $i' = j = I$.

$$\begin{aligned} H_1 &= V_{I,II;I,II}\ c_{I,\uparrow}^\dagger c_{II,\uparrow}^\dagger c_{II,\uparrow} c_{I,\uparrow} + V_{I,II;II,I}\ c_{II,\uparrow}^\dagger c_{I,\uparrow}^\dagger c_{II,\uparrow} c_{I,\uparrow} = \\ &= (I_{\text{direct}} - I_{\text{exchange}}) c_{I,\uparrow}^\dagger c_{II,\uparrow}^\dagger c_{II,\uparrow} c_{I,\uparrow} \end{aligned} \tag{8.57}$$

The matrix elements $V_{II,I;II,I} = V_{I,II;I,II}$ and $V_{II,I;II,I} = V_{I,II;I,II}$ are I_{direct} and I_{exchange}, as can be seen by comparing Eqs. (8.12), (8.13) and (8.34). The energy correction is then $\langle 0|c_{II,\sigma} c_{I,\sigma}|H_1|c_{I,\sigma}^\dagger c_{II,\sigma}^\dagger|0\rangle = I_{\text{direct}} - I_{\text{exchange}}$.

In the $S^z = 0$ channel, the restriction that only $V_{\sigma,\sigma';\sigma,\sigma'} \neq 0$ leaves terms

$$H_1 |c_{I,\uparrow}^\dagger c_{II,\downarrow}^\dagger|0\rangle = \Big(V_{II,I;II,I} c_{II,\downarrow}^\dagger c_{I,\uparrow}^\dagger c_{I,\uparrow} c_{II,\downarrow} + V_{I,II;II,I} c_{I,\downarrow}^\dagger c_{II,\uparrow}^\dagger c_{I,\uparrow} c_{II,\downarrow}\Big)|c_{I,\uparrow}^\dagger c_{II,\downarrow}^\dagger|0\rangle =$$

$$= V_{II,I;II,I}|c^\dagger_{I,\uparrow}c^\dagger_{II,\downarrow}|0\rangle - V_{II,I;I,II}|c^\dagger_{I,\downarrow}c^\dagger_{II,\uparrow}|0\rangle \ .$$

In the subspace spanned by $|c^\dagger_{I,\downarrow}c^\dagger_{II,\uparrow}|0\rangle$ and $|c^\dagger_{I,\uparrow}c^\dagger_{II,\downarrow}|0\rangle$, the electron-electron interaction is a 2×2 matrix

$$H_1 = \begin{pmatrix} I_{\text{direct}} & -I_{\text{exchange}} \\ -I_{\text{exchange}} & I_{\text{direct}} \end{pmatrix} \tag{8.58}$$

whose lower-energy eigenstate is $\frac{1}{\sqrt{2}}(|c^\dagger_{I,\downarrow}c^\dagger_{II,\uparrow}|0\rangle + |c^\dagger_{I,\uparrow}c^\dagger_{II,\downarrow}|0\rangle)$, the third triplet.

8.3.2 Electrons in a solid

Nolting or Schwabl

The formalism with creators/annihilators only really shows its advantage for more than two particles, e.g., for the electrons in a solid.

A simple model is "jellium", where the ionic background is assumed to be homogeneous. The electrons thus only see a positively charged background rather than a periodic lattice potential, and they are confined to a box with volume $V = L_x \times L_y \times L_z$. The Hamiltonian is

$$H = H_{\text{ions}} + H_{\text{ion-el.}} + H_{\text{kin, el.}} + H_{\text{el.-el.}} \ . \tag{8.59}$$

where the part describing ions, $H_{\text{ions}} + H_{\text{ion-electron}}$ is of course vastly simplified, as ions are only a smooth background. The electronic kinetic energy $H_{\text{kin, el.}}$ is simply the kinetic energy $\frac{p^2}{2m}$ of electrons confined to a box and $H_{\text{el.-el.}}$ is the electron-electron Coulomb interaction. If there was just one electron in the solid, its eigenstates would be given by plane waves

$$\psi_{\vec{k}}(\vec{r}) = \frac{1}{\sqrt{V}} \mathrm{e}^{i\vec{k}\vec{r}} \ , \tag{8.60}$$

as the problem then reduces to one particle in a box with infinite walls. In addition, the spin has to be taken into account and as it is not coupled to position space in this Hamiltonian, the full wave function $|k,\sigma\rangle = |k\rangle \otimes |\sigma\rangle$. Replacing infinite walls by periodic boundary conditions – a more common choice in condensed-matter theory – does not change this, the allowed momenta are then quantized to $\vec{k} = (k_x, k_y, x_z) = 2\pi \cdot (i_x/L_x, i_y/Ly, i_z/Lz) - \pi$. [7]

As one-particle eigenstates are plane waves, we also use them to define creation (annihilation) operators $a^\dagger_{k,\sigma}$ ($a_{k,\sigma}$), which create (annihilate) an electron with momentum k and spin σ. Following Sec. 8.2.1, the kinetic energy must be of the form

$$H_{\text{kin, el.}} = \sum_{\substack{k,k' \\ \sigma,\sigma'}} e_{\substack{k,k' \\ \sigma,\sigma'}} a^\dagger_{k',\sigma'} a_{k,\sigma} \ . \tag{8.61}$$

Including spin space, the one-particle kinetic energy is $H^1_{\text{kin, el.}} = \frac{p^2}{2m} \otimes \mathbb{I}$, i.e. spin is not modified. The matrix element is

$$e_{\substack{k,k' \\ \sigma,\sigma'}} = \langle k,\sigma|\tfrac{p^2}{2m}|k',\sigma'\rangle = \int d^3r \ \psi^*_{k'}(\vec{r}) \ \frac{-\hbar^2\nabla^2}{2m} \ \psi_k(\vec{r}) \cdot \underbrace{\langle\sigma'|\sigma\rangle}_{=\delta_{\sigma,\sigma'}} =$$

[7] Quantization also arises for infinite walls.

$$= \frac{-\hbar^2}{2m}\frac{\delta_{\sigma,\sigma'}}{V}\int d^3r\ \mathrm{e}^{-i\vec{k}'\vec{r}}\underbrace{\nabla^2\mathrm{e}^{-i\vec{k}\vec{r}}}_{k^2\mathrm{e}^{-i\vec{k}\vec{r}}} = \frac{-\hbar^2k^2}{2m}\delta_{\sigma,\sigma'}\delta_{k,k'}\ . \tag{8.62}$$

The kinetic energy consequently conserves momentum – a result of translational invariance within the solid. The many-body kinetic energy is then

$$H_{\text{kin, el.}} = -\sum_k \frac{\hbar^2k^2}{2m}\sum_\sigma a^\dagger_{k,\sigma}a_{k,\sigma} = -\sum_k \frac{\hbar^2k^2}{2m}\sum_\sigma n_{k,\sigma}\ . \tag{8.63}$$

The ground state for this part of the Hamiltonian alone is found by filling the lowest $N/2$ states with 2 electrons each, as the energy depends here only on $|k|$, this is a sphere in k-space, the "Fermi sphere".

The combined ion-ion and ion-electron interactions actually diverge if one assumes a Coulomb potential. Taken by themselves, the ion-electron interaction would be too strong, which should – if taken seriously – tend to collapse the solid. However, these terms are exactly balanced by a repulsive and equally divergent term in the electron-electron interaction, stabilizing the solid, see Sec 8.3.2.1 below.

We focus now on the rest of the electron-electron interaction, where we introduce the Fourier transform

$$\hat{V} = V(\vec{r}_1-\vec{r}_2) = \frac{e^2}{4\pi\epsilon_0}\frac{1}{|\vec{r}_1-\vec{r}_2|} = \frac{1}{V}\sum_q \mathrm{e}^{i\vec{q}(\vec{r}_1-\vec{r}_2)}V_q = \frac{e^2}{\epsilon_0}\sum_q \mathrm{e}^{i\vec{q}(\vec{r}_1-\vec{r}_2)}\frac{1}{q^2}\ . \tag{8.64}$$

The $q=0$ term is the one exactly cancelled by the background, the sum is consequently taken to run over $\vec{q}\neq 0$ only in the following.

The matrix element for the interaction can be evaluated as

$$\begin{aligned}
&{}^1\langle k_1,\sigma_1|\,{}^2\langle k_2,\sigma_2|\hat{V}|k_3,\sigma_3\rangle^1|k_4,\sigma_4\rangle^2 = \qquad (8.65)\\
&= {}^1\langle\sigma_1|\sigma_3\rangle^1\,{}^2\langle\sigma_2|\sigma_4\rangle^2\cdot\frac{1}{V^2}\int d^3r_1\int d^3r_2\ \mathrm{e}^{-i\vec{k}_1\vec{r}_1}\mathrm{e}^{-i\vec{k}_2\vec{r}_2}V(\vec{r}_1-\vec{r}_2)\mathrm{e}^{i\vec{k}_3\vec{r}_1}\mathrm{e}^{i\vec{k}_4\vec{r}_2} =\\
&= \delta_{\sigma_1,\sigma_3}\delta_{\sigma_2,\sigma_4}\frac{1}{V^3}\sum_{q\neq 0}\int d^3r_1\int d^3r_2\ \mathrm{e}^{-i\vec{k}_1\vec{r}_1}\mathrm{e}^{-i\vec{k}_2\vec{r}_2}\mathrm{e}^{-i\vec{q}(\vec{r}_1-\vec{r}_2)}V_q\mathrm{e}^{i\vec{k}_3\vec{r}_1}\mathrm{e}^{i\vec{k}_4\vec{r}_2} =\\
&= \delta_{\sigma_1,\sigma_3}\delta_{\sigma_2,\sigma_4}\frac{1}{V^3}\sum_{q\neq 0}V_q\underbrace{\int d^3r_1\mathrm{e}^{-i(\vec{k}_1-\vec{k}_3+\vec{q})\vec{r}_1}}_{=V\delta_{\vec{k}_1+\vec{q},\vec{k}_3}}\underbrace{\int d^3r_2\ \mathrm{e}^{-i(\vec{k}_2-\vec{k}_4-\vec{q})\vec{r}_2}}_{=V\delta_{\vec{k}_2-\vec{q},\vec{k}_4}} =\\
&= \frac{\delta_{\sigma_1,\sigma_3}\delta_{\sigma_2,\sigma_4}}{V}\sum_{q\neq 0}\frac{e^2}{\epsilon_0q^2}\delta_{\vec{k}_1,\vec{k}_3-\vec{q}}\delta_{\vec{k}_2,\vec{k}_4+\vec{q}}\ .
\end{aligned}$$

The interaction in second quantization is then, see (8.54):

$$H_{\text{el.-el.}} = \frac{1}{2V}\sum_{\substack{q\neq 0;p,k\\ \sigma,\sigma'}}\frac{e^2}{\epsilon_0q^2}\ a^\dagger_{k-q,\sigma}a^\dagger_{p+q,\sigma'}a_{p,\sigma'}a_{k,\sigma}\ . \tag{8.66}$$

Applied to the ground state of the kinetic energy, the Fermi sphere, this operator takes two particles with momentum k and p (with $|k|,|p|\leq k_{\text{Fermi}}$) and moves them to $k-q$ and

$p+q$: The total momentum consequently remains the same, again a result of translational invariance.

In first-order perturbation theory, only $k = p + q$ and $p = k - q$ contribute, which is the exchange term. (The direct term with $q = 0$ was explicitly taken out to compensate the background.) In higher orders, states $k - q$ and $p + q$ can be outside the Fermi sphere, such contributions are more important if both the original and the final states are close to the Fermi surface, because the process then has to pay less kinetic energy.

8.3.2.1 Compensation of the $q = 0$ term with the ionic background

In this subsection, we shortly discuss the above mentioned divergent terms of the Coulomb interaction. For $q = 0$, the Fourier transform $\int d^3r \frac{1}{|r|}$ diverges unless a factor $\mathrm{e}^{-\alpha|r|}$ is introduced, where we would aim to let α go to 0 in the end again. The $q = 0$ term of the electron-electron interaction then becomes

$$V_{k,p;k,p} = \frac{e^2}{V\epsilon_0}\frac{1}{q^2+\alpha^2} = \frac{e^2}{V\epsilon_0}\frac{1}{\alpha^2} \tag{8.67}$$

and hence

$$\begin{aligned}
\frac{e^2}{2V\epsilon_0\alpha^2}\sum_{\substack{p,k\\ \sigma,\sigma'}} a^\dagger_{k,\sigma}a^\dagger_{p,\sigma'}a_{p,\sigma'}a_{k,\sigma} &= -\frac{e^2}{2V\epsilon_0\alpha^2}\sum_{\substack{p,k\\ \sigma,\sigma'}} a^\dagger_{k,\sigma}\underbrace{a^\dagger_{p,\sigma'}a_{k,\sigma}}_{=\delta_{k,p}\delta_{\sigma,\sigma'}-a_{k,\sigma}a^\dagger_{p,\sigma'}} a_{p,\sigma'} = \\
&= \frac{e^2}{2V\epsilon_0\alpha^2}\sum_{\substack{p,k\\ \sigma,\sigma'}} a^\dagger_{k,\sigma}a_{k,\sigma}a^\dagger_{p,\sigma'}a_{p,\sigma'} - \frac{e^2}{2\epsilon_0\alpha^2}\sum_{\substack{p,k\\ \sigma,\sigma'}} a^\dagger_{k,\sigma}\delta_{k,p}\delta_{\sigma,\sigma'}a_{p,\sigma'} = \\
&= \frac{e^2}{2V\epsilon_0\alpha^2}\left(\sum_{k,\sigma}\hat{n}_{k,\sigma}\right)\left(\sum_{p,\sigma'}\hat{n}_{p,\sigma'}\right) - \frac{e^2}{2V\epsilon_0\alpha^2}\sum_{k,\sigma}\hat{n}_{k,\sigma}\ .
\end{aligned} \tag{8.68}$$

The sum over all occupation numbers must give the total number of electrons N and with the average density $\bar{n} = N/V$, the expectation value for the interaction energy *per particle* becomes

$$\frac{e^2}{2\epsilon_0\alpha^2}\bar{n} - \frac{e^2}{2V\epsilon_0\alpha^2}\ . \tag{8.69}$$

We are here interested in the limit $\alpha \to 0$, but also in the so-called "thermodynamic limit", i.e., an infinitely large system where $N \to \infty$ and $V \to \infty$, but the average density $\bar{n} = \frac{N}{V}$ remains constant. If we take the thermodynamic limit first, the second term vanishes, but the first remains finite and then diverges for the limit $\alpha \to 0$.

Let us now look at the Coulomb energy of the ionic background. In the jellium model, the ions are treated as a homogeneous, and in a way classical, background of density $\bar{n}$, whose and charge $e\bar{n}$ exactly cancels the electronic charge $-e\bar{n}$ to ensure charge neutrality. Its Coulomb energy is

$$\frac{e^2}{4\pi\epsilon_0}\frac{1}{2}\int d^3r_1\int d^3r_2\frac{n(\vec{r}_1)n(\vec{r}_2)}{|\vec{r}_1-\vec{r}_2|} = \frac{e^2}{4\pi\epsilon_0}\frac{1}{2}\int d^3r_1\int d^3r_2\frac{\bar{n}\bar{n}}{|\vec{r}_1-\vec{r}_2|} \tag{8.70}$$

where the factor $\frac{1}{2}$ makes up for the fact that every pair of $(\vec{r}, \vec{r}')$ comes up twice in the integral. Again using the convergence factor $\mathrm{e}^{-\alpha(\vec{r}_1-\vec{r}_2)}$, it becomes

$$\frac{e^2}{2\epsilon_0}\frac{1}{\alpha^2}\bar{n}^2 V \tag{8.71}$$

and the energy per particle is identical to the electronic interaction. Finally, we need to take into account the interaction fo the ionic background and the electrons. The opposite charges make this term attractive and each electron fells a constant potential

$$\frac{-e^2}{4\pi\epsilon_0}\int d^3r \frac{\bar{n}}{|\vec{r}|} = \frac{-e^2}{\epsilon_0\alpha^2}\,. \tag{8.72}$$

Multiplied by the number N of electrons, this term is exactly the sum of the two previous terms[8] and all terms taken together vanish for any α.

8.3.2.2 Various representations

Nolting

An often chosen basis for the definition of the creation and annihilation operators are the eigenstates of the one-particle part $H_{\text{ion-el.}} + H_{\text{kin, el.}}$ of the electronic hamiltonian. In the example of jellium, these are plane waves with discrete wave numbers, eigenstates of the kinetic energy. Even if the ionic background is modelled in a more sophisticated way than as a homogeneous background, its impact on electrons can still (usually) be modelled as a time-independent potential, i.e., $H_{\text{ion-el.}}$ is a one-particle term. Due to lattice periodicity with respect to lattice vectors $\vec{R}$, Bloch's theorem ensures that the eigenstates of the sum of kinetic energy and the potential, $H_{\text{ion-el.}} + H_{\text{kin, el.}}$, have the form

$$\psi_k(\vec{r}+\vec{R}) = \mathrm{e}^{ikR}\psi_k(\vec{r}), \quad \text{resp.} \quad \psi_k(\vec{r}) = \mathrm{e}^{ikr}u_k(\vec{r}) \tag{8.73}$$

with a lattice-periodic $u_k(\vec{r}) = u_k(\vec{r}+\vec{R})$. Creators and annihilators are then introduced for these Bloch states and the one-particle part of the Hamiltonian becomes

$$H_{\text{ion-el.}} + H_{\text{kin, el.}} = \sum_{\vec{j},\sigma} \epsilon(\vec{k}) a^\dagger_{\vec{k},\sigma} a_{\vec{k},\sigma}\,, \tag{8.74}$$

with the band structure $\epsilon(\vec{k})$.

Electronic operators can be transformed into position space in the same way as k-dependent phonon and photon operators were in Eqs. (7.27) and (7.55). The resulting

$$\psi^\dagger_\sigma(\vec{r}) = \frac{1}{\sqrt{N}}\sum_{\vec{k}} \mathrm{e}^{ikr} a^\dagger_{\vec{k},\sigma} \quad \text{and} \quad \psi_\sigma(\vec{r}) = \frac{1}{\sqrt{N}}\sum_{\vec{k}} \mathrm{e}^{-ikr} a_{\vec{k},\sigma} \tag{8.75}$$

are then electronic "field operators".

[8] Another way to see this is to note that the interaction between a homogeneous ion and a homogeneous electron density follows the same considerations as the ion-ion interaction, but there is no factor $\frac{1}{2}$, because "electron at $\vec{r}$ and ion at $\vec{r}'$" and "electron at $\vec{r}'$ and ion at $\vec{r}$" both contribute.

Another often used position-space representation are Wannier functions

$$\psi_{\vec{R}}(\vec{r}) = \frac{1}{\sqrt{N}} \sum_{\vec{k}} \mathrm{e}^{-ikR} \psi_{\vec{k}}(\vec{r}) \tag{8.76}$$

indexed with a lattice position $\vec{R}$. The corresponding operators

$$a^\dagger_{\vec{R},\sigma} = \frac{1}{\sqrt{N}} \sum_{\vec{k}} \mathrm{e}^{-ikR} a^\dagger_{\vec{k},\sigma} \tag{8.77}$$

refer to electrons well localized around lattice site $\vec{R}$, because $\psi_{\vec{R},\sigma}(\vec{r}) \approx 0$ for large $|\vec{R}-\vec{r}| \gg a$. As an illustration, let us look at a very simple form of the periodic function $u_k(\vec{r})$ an the corresponding Bloch and Wannier states:

$$u_k(\vec{r}) = \tfrac{1}{\sqrt{N}} \sum_{\vec{R}} \delta(\vec{r}-\vec{R}) \quad \Rightarrow \quad \psi_k(\vec{r}) = \mathrm{e}^{ikr} u_k(\vec{r}) = \tfrac{1}{\sqrt{N}} \sum_{\vec{R}} \mathrm{e}^{ikR} \delta(\vec{r}-\vec{R}) \tag{8.78}$$

$$\psi_{\vec{R}}(\vec{r}) = \frac{1}{N} \sum_{\vec{k}} \mathrm{e}^{-ikR} \sum_{\vec{R}'} \mathrm{e}^{ikR'} \delta(\vec{r}-\vec{R}') = \frac{1}{N} \sum_{\vec{R}'} \delta(\vec{r}-\vec{R}') \underbrace{\sum_{\vec{k}} \mathrm{e}^{-ikR+ikR'}}_{=N\delta(\vec{R}-\vec{R}')} = \delta(\vec{r}-\vec{R}) \,, \tag{8.79}$$

i.e., the Wannier function $\psi_{\vec{R}}(\vec{r})$ picks out the part localized at $\vec{R}$ from the lattice periodic $u_k(\vec{r})$.

9 Relativistic Quantum Mechanics

To a large extent Schwabl, some Sakurai II

The theory of special relativity was introduced to make electrodynamics and mechanics consistent. The main issue is that a constant speed of light implies natural laws should be invariant under Lorentz transformations rather than obeying Galilean invariance. As we are going to see, this leads to solutions with negative energy. As they are hard to understand for a single particle, one then almost necessarily ends up with a many-body theory, i.e., a field theory based on operators analogous to (8.75).

9.1 Relativistic Mechanics, Four-vectors, ...

The notation commonly used in discussion special relativity will shortly be reviewed here. The most striking novelty of special relativity, resp. of Lorentz transformations, is that time and space can to some extent transform into each other. Consequently, they have to be treated together and are combined into a "four-vector" with one temporal and three spatial dimensions. The "contra-variant" vector

$$x^\mu = \left(x^0 = ct, x^1 = x, x^2 = y, x^3 = z\right) = (ct, \vec{x}) \tag{9.1}$$

is written with a raised Greek index, the "co-variant"

$$x_\mu = (x_0, x_1 = -x, x_2 = -y, x_3 = -z)\,(ct, -\vec{x}) \tag{9.2}$$

with a lowered index. To get from the co- to the contra-variant form of a four-vector and back, the metric tensor is introduced

$$x_\mu = g_{\mu,\nu}x^\nu \quad \text{and} \quad x^\mu = g^{\mu,\nu}x_\nu \quad \text{with} \quad g_{\mu\nu} = g^{\mu\nu} = \begin{cases} 1 & \text{for} \quad \mu = \nu = 0 \\ -1 & \text{for} \quad \mu = \nu \neq 0 \\ 0 & \text{for} \quad \mu \neq \nu \end{cases} \tag{9.3}$$

(In expression with a repeated index – one lowered and one raised – this index should be summed over.) An important thing to note is the position of the index in the four-gradient:

$$\partial_\mu = \frac{\partial}{\partial x^\mu} = (\frac{\partial}{\partial ct}, \frac{\partial}{\partial x^k}) = (\frac{1}{c}\partial_t, \nabla)\ , \tag{9.4}$$

i.e., ∂_μ with a lowered index refers to the derivative with respect to x^μ with a raised index. [For the reason seeEq. (9.12).]

The gist of special relativity is captured in these rules of how to deal with "four vectors". Particularly important, the inner product of two such vectors

$$x \cdot y = x^\mu y_\mu = \sum_\mu x_\mu y^\mu = x^0 y^0 - \vec{x}\vec{y} = c^2 t_x t_y - \vec{x}\vec{y}\ . \tag{9.5}$$

is indeed a *scalar*, i.e., invariant under Lorentz transformations and independent of the reference frame. In particular, the norm of a vector

$$x \cdot x = x_\mu x^\mu = c^2 t^2 - \|\vec{x}\|^2 \tag{9.6}$$

is invariant – in contrast to $\|\vec{x}\|^2$ by itself, which is only invariant under non-relativistic Galilei transformations.

The trajectory $x = (ct, \vec{x})$ of a particle is also referred to as its *world line* x. Since Lorentz transformations also affect time, the derivative with respect to $t = \frac{x^0}{c}$ depends on the reference frame. A uniquely defined reference frame of particular importance for a particle is its rest frame $(c\tau, \vec{x} = 0)$, where the time component is its proper time. Due to the invariance of the norm, it has to be

$$c^2\tau^2 - 0^2 = c^2 t^2 - \vec{x}^2 \quad \Rightarrow \quad d\tau = \sqrt{1 - \frac{\vec{v}^2}{c^2}}\, dt \; . \tag{9.7}$$

(Only if the particle moves with a uniform velocity, i.e., $\vec{x} = \vec{v}t$, do we also find $\tau = t\sqrt{1 - \frac{\vec{v}^2}{c^2}}$.) The derivative of the world line $x = (ct, \vec{v}\, t)$ with respect to the proper time is

$$\frac{\partial x}{\partial \tau} = \frac{\partial x}{\partial t}\frac{dt}{d\tau} = \frac{1}{\sqrt{1 - \frac{\vec{v}^2}{c^2}}} \begin{pmatrix} c \\ \vec{v} \end{pmatrix} \; . \tag{9.8}$$

As $\frac{dx}{d\tau}$ is defined using the unique proper time, it is used to define momentum p with components

$$p^\mu = m\frac{\partial x^\mu}{\partial \tau} = \frac{1}{\sqrt{1 - \frac{\vec{v}^2}{c^2}}} \begin{pmatrix} mc \\ m\vec{v} \end{pmatrix} = \begin{pmatrix} E/c \\ \vec{p} \end{pmatrix} \; . \tag{9.9}$$

The zeroth component of the four-momentum can be taken as the energy. As the norm of a four-vector is the same in any reference system, it always has to be equal to that found in the rest frame, where $\vec{p} = 0$ and $E = mc^2$, yielding

$$p_\mu p^\mu = \frac{E^2}{c^2} - \vec{p}^2 = m^2 c^2 \; . \tag{9.10}$$

9.1.1 Lorentz transformations

The idea of special relativity is that the laws of physics should be the same in all inertial frames. As one consequence, the d'Alembert operator has to be invariant under Lorentz transformations. First, some comments on the type of transformations that are relevant here: A particle not feeling any forces moves with constant velocity and its world line $(ct, \vec{x})$ is a straight line in four-dimensional space time. This qualitative aspect should also hold in other equivalent reference frames, and straight lines remain straight lines under linear transformations. We consequently expect linear transformations of vector, i.e. matrix multiplications:

$$x' = \Lambda x + a \quad \text{resp} \quad x'^\nu = \Lambda^\nu{}_\mu x^\mu + a^\nu \tag{9.11}$$

a is a translation in space-time, which we will not discuss much here, but which is allowed. The currently more interesting part is captured by the matrix Λ, which one can see as a kind of generalized rotation in four-dimensional space time. It forms a group that includes rotations of three-dimensional space as a subgroup.

Derivatives with respect to x' can be expressed in terms of those with respect to x using the chain rule

$$\partial_\mu = \frac{\partial}{\partial x^\mu} = \frac{\partial}{\partial x'^\nu} \underbrace{\frac{\partial x'^\nu}{\partial x^\mu}}_{=\Lambda^\nu{}_\mu} = \Lambda^\nu{}_\mu \partial'_\nu \; . \tag{9.12}$$

One sees here that $\frac{\partial}{\partial^\mu}$ transforms like a co-variant vector, which is why it is denoted as ∂_μ. The d'Alembert operator then transforms as

$$\partial_\mu \partial^\mu = \partial_\mu g^{\mu\nu} \partial_\nu = \Lambda^\rho{}_\mu \partial'_\rho g^{\mu\nu} \Lambda^\sigma{}_\nu \partial'_\sigma = \partial'_\rho \partial'^\rho = \partial'_\rho g^{\rho\sigma} \partial'_\sigma \tag{9.13}$$

which requires Λ to fulfill

$$\Lambda^\rho{}_\mu g^{\mu\nu} \Lambda^\sigma{}_\nu = g^{\rho\sigma} \quad \text{resp.} \quad \Lambda g \Lambda^T = g \; . \tag{9.14}$$

This equation can be seen as an analogue of $U^\dagger U = \mathbb{I}$ required for unitary transformations encoding symmetries in Hilbert spaces: scalar products (and hence norms) have to be left invariant in both cases.

9.2 Klein-Gordon Equation

The first attempt to find a relativistic wave equation start from the above relation for the norm of the four momentum and then uses the same "canonical quantization" that was successful in the non-relativistic case: $E \to i\hbar\partial_t$ and $\vec{p} \to -i\hbar\nabla$. The quantized Eq. (9.10) becomes then the Klein-Gordon equation

$$\left(-\frac{\hbar^2}{c} \frac{\partial^2}{\partial t^2} + \hbar^2 \nabla^2 \right) \psi(t, \vec{x}) = m^2 c^2 \psi(t, \vec{x}) \quad \text{or} \tag{9.15}$$

$$\left(\underbrace{\frac{1}{c^2} \frac{\partial^2}{\partial t^2} - \Delta}_{=\partial_\mu \partial^\mu = \Box} + \underbrace{\left(\frac{mc}{\hbar} \right)^2}_{=\lambda^2} \right) \psi(t, \vec{x}) = 0 \; . \tag{9.16}$$

The equation is relativistically invariant, because the d'Alembert operator $\Box = \partial_\mu \partial^\mu$ is. In fact, the observation that light, whose wave equation is described by this operator, is invariant under Lorentz transformations was the driving force behind the development of special relativity. The rest of the wave equation is then just a product of constants, hence also invariant.

Free-particle solutions of the Klein-Gordon equation are plane waves, apart from a normalization, they are given by

$$\psi(t, \vec{x}) = \mathrm{e}^{\frac{i}{\hbar}(Et - \vec{p}\vec{x})} \tag{9.17}$$

with $E = \pm\sqrt{\vec{p}^2c^2 + m^2c^4}$. This looks extremely similar to the free-particle solutions (1.37) of the non-relativistic Schrödinger equation. However, as discussed in Sec. 1.3, the non-relativistic energy-momentum relation $E = \frac{p^2}{2m}$ enforces $E \geq 0$, because wave functions for imaginary p would diverge. The situation is different here, where the negative sign for E can go together with real $\vec{p}$ and thus still leads to a normalizable plane-wave solution. As we are going to see, these negative energies are a recurring feature of relativistic wave equations.

The negative-energy solutions together with the absence of a continuity equation for any quantity that can be seen as a probability distribution originally led to a dismissal of the Klein-Gordon equation. It is, however, a perfectly valid relativistic wave equation, just not for electrons. The continuity equation can be interpreted in terms of a charge density rather than a probability density and leads to a many-body situation. Ending up with many-body situations even when trying to discuss a single particle is a recurrent aspect of relativistic quantum mechanics, which can be seen as a result fo the equivalence of matter and energy, where particles can be created and annihilated.

9.2.1 Continuity equation

From the non-relativistic Schrödinger equation, a continuity equation can be obtained that relates a change in density to a current. As this density has to be ≥ 0, it can be interpreted as a probability density, and the current accordingly as a probability current. In the case of the Klein-Gordon equation, the corresponding steps are:

$$\begin{aligned}
\psi^*(t,\vec{x}) \cdot \left(\frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \Delta + \left(\frac{mc}{\hbar}\right)^2\right)\psi(t,\vec{x}) &= 0 \\
\psi(t,\vec{x}) \cdot \left(\frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \Delta + \left(\frac{mc}{\hbar}\right)^2\right)\psi^*(t,\vec{x}) &= 0 \\
\frac{1}{c^2}\frac{\partial}{\partial t}\left(\psi^*\frac{\partial\psi}{\partial t} - \frac{\partial\psi^*}{\partial t}\psi\right) - \nabla\left(\psi^*\nabla\psi - (\nabla\psi^*)\psi\right) &= 0
\end{aligned} \tag{9.18}$$

While this has the form of a continuity equation $\partial_t\rho - \mathrm{div}\vec{j} = 0$, the density

$$\rho(t,\vec{x}) = \frac{1}{c^2}\left(\psi^*\frac{\partial\psi}{\partial t} - \frac{\partial\psi^*}{\partial t}\psi\right) \tag{9.19}$$

does not have to be positive. As the Klein-Gordon equation is second order in time, initial conditions need to specify $\partial_t\psi$ in addition to ψ at time $t = 0$ and this freedom permits negative ρ for any choice of $\psi(t = 0, \vec{x})$.

Since ρ can become smaller than 0, it cannot be the probability density describing a quantum particle, which makes the Klein-Gordon equation difficult to interpret for a single particle. It turned out later that ρ is a valid *charge* density in a many-body situation with both positive and negative charges.

9.3 Dirac Equation

A second attempt at a relativistic wave equation originally sought to fix the "problems" with the Klein-Gordon equation. While this did not work out as hoped, the Dirac equation turned out to describe particles with spin 1/2, e.g., electrons.

9.3.1 Motivation as a first-order differential equation

The idea was to obtain an equation first order in time derivatives, because the second order leads to the issues with the probability density in the case of the Klein-Gordon equation. Due to the required relativistic invariance, the equation would then also have to be first order in spatial derivatives. For simple numbers, relativistic invariance cannot be achieved with a first-order equations. One thus has to assume that the equation has vector valued solutions and is a matrix equation. As we also want a continuity equation for a real and positive probability density, the matrix equation should be hermitian, as we are going to see below. Finally, since the energy-momentum relation giving the Klein-Gordon equation is very reasonable for relativistic particles, we would like the components of the solution vector to fulfill it and thus the Klein-Gordon equation.

These considerations lead to the ansatz

$$i\hbar\frac{\partial}{\partial t}\psi(t,\vec{x}) = \left(-i\hbar c\alpha^k\partial_k + \beta mc^2\right)\psi(t,\vec{x})\ , \tag{9.20}$$

where β and α^k are matrices whose dimension and properties remain to be determined.

If each component ist to fulfill the Klein-Gordon equation, we need the second derivative with respect to time. Since taking the derivative is – for the solution – supposed to give the same as applying the right-hand side of (9.20), we aply it twice. This will obviously yield second-order spatial derivatives that can be compared to the Klein-Gordon equation. The idea is that the Klein-Gordon equation for a single wave function might not be a good description of an electron, but a vector of several wave functions, where each component fulfills it, might be.

The squared differential operators leads to

$$-\frac{\partial^2}{\partial t^2}\psi(t,\vec{x}) = \left[-c^2\sum_{k,l}\alpha^k\partial_k\alpha^l\partial_l + \beta^2\left(\frac{mc^2}{\hbar}\right)^2 - \frac{2imc^3}{\hbar}\sum_k(\alpha^k\beta+\beta\alpha^k)\partial_k\right]\psi(t,\vec{x})\ ,$$

to be compared to (9.21)

$$-\frac{\partial^2}{\partial t^2}\psi(t,\vec{x}) = \left[-c^2\nabla^2\psi(t,\vec{x}) + \left(\frac{mc^2}{\hbar}\right)^2\right]\psi(t,\vec{x})\ . \tag{9.22}$$

As the partial derivatives commute (and can be switched with the matrices), the first double sum on the right can be symmetrized to $\frac{1}{2}\sum_{k,l}(\alpha^k\alpha^l + \alpha^l\alpha^k)\partial_k\partial_l$. As this part of the equations should become $-\frac{\vec{p}^2}{\hbar^2} = \Delta$, we can conclude

$$\frac{1}{2}(\alpha^k\alpha^l + \alpha^l\alpha^k) = \delta_{k,l}\mathbb{I}\ . \tag{9.23}$$

The last term on the right should vanish, because the Klein-Gordon equation does not contain forst-order spatial derivatives, giving

$$\beta\alpha^k + \alpha^k\beta = 0 \tag{9.24}$$

and finally, the mass term yields

$$\beta^2 = \mathbb{I}\ . \tag{9.25}$$

A continuity equation should be a relation between first-order derivatives of quantities that can be identified as density and current. It is obtained by taking the complex conjugate of the Dirac equation, multiplying it with the wave function and subtracting it from the equation itself multiplied by $\psi^\dagger$, which is the vector with complex conjugate components $\psi^\dagger = (\psi_1^*, \psi_2^*, \dots)$.

$$\begin{aligned}
&\psi^\dagger(t,\vec{x}) i\hbar \frac{\partial}{\partial t}\psi(t,\vec{x}) - \psi(t,\vec{x}) \left(-i\hbar \frac{\partial}{\partial t}\psi^\dagger(t,\vec{x})\right) = \\
&\quad = \psi^\dagger(t,\vec{x}) \left(-i\hbar c \alpha^k \partial_k + \beta m c^2\right) \psi(t,\vec{x}) - \psi(t,\vec{x}) \left(i\hbar c \big(\alpha^k\big)^\dagger \partial_k + \beta^\dagger m c^2\right) \psi^\dagger(t,\vec{x}) \\
&i\hbar \left(\psi^\dagger(t,\vec{x}) \frac{\partial}{\partial t}\psi(t,\vec{x}) + \psi(t,\vec{x}) \frac{\partial}{\partial t}\psi^\dagger(t,\vec{x})\right) = \\
&\quad = -i\hbar c \left(\psi^\dagger(t,\vec{x}) \alpha^k \partial_k \psi(t,\vec{x}) + \psi(t,\vec{x}) \big(\alpha^k\big)^\dagger \partial_k \psi^\dagger(t,\vec{x})\right) + (\beta - \beta^\dagger) m c^2 \psi^\dagger(t,\vec{x}) \psi(t,\vec{x})
\end{aligned}$$

For this equation to become an expression in terms of "derivatives of something", the last term has to vanish, i.e. $\beta = \beta^\dagger$. Further setting $\big(\alpha^k\big)^\dagger = \alpha^k$ gives

$$\frac{\partial}{\partial t}\left(\underbrace{\psi^\dagger(t,\vec{x})\psi(t,\vec{x})}_{=\rho}\right) = -c\partial_k \left(\underbrace{\psi^\dagger(t,\vec{x})\alpha^k\psi(t,\vec{x})}_{=j^k/c}\right) , \tag{9.26}$$

where a density $\rho \geq 0$ and a current j can be defined. With $j^0 = c\rho$, the continuity equation can be written as an equation for a Lorentz-scalar $\partial_\mu j^\mu = 0$. The density can here be interpreted as a probability density. In order to achieve this, β and α^k have to be hermitian.

The matrices α^k and β have turn out to be hermitian and to anti-commute, moreover, their square is $\mathbb{I}$. All eigenvalues of the squares are consequently 1 and the eigenvalues of the matrices themselves have to be ± 1, because hermitian matrices have real eigenvalues. Moreover, the cyclic invariance of the trace together with (9.24), (9.23) and (9.25) can be used to find that their traces vanish:

$$\begin{aligned}
\alpha^k &= \beta\beta\alpha^k = -\beta\alpha^k\beta \\
\text{tr}\ \alpha^k &= -\text{tr}\ \beta\alpha^k\beta = -\text{tr}\ \alpha^k\beta\beta = -\text{tr}\ \alpha^k
\end{aligned} \tag{9.27}$$

The number of eigenvalues $+1$ and that of -1 must consequently be the same and the dimension of the matrices must be even. A complete basis of the 2×2 traceless hermitian matrices are the three Pauli matrices, but we need four distinct matrices. The dimension is therefore larger than 2; 4 turns out to work. One representation (written in terms of 2×2 matrices) is

$$\beta = \begin{pmatrix} \mathbb{I} & 0 \\ 0 & -\mathbb{I} \end{pmatrix}, \quad \alpha^k = \begin{pmatrix} 0 & \sigma^k \\ \sigma^k & 0 \end{pmatrix} , \tag{9.28}$$

where σ^k are the Pauli matrices (3.11).

A formulation that more clearly expresses relativistic invariance is found by multiplying the Dirac equation by $\beta/(\hbar c)$:

$$i\underbrace{\beta}_{\gamma^0}\frac{1}{c}\frac{\partial}{\partial t}\psi(t,\vec{x}) = \left(-i\underbrace{\beta\alpha^k}_{\gamma^k}\partial_k + \frac{mc}{\hbar}\right)\psi(t,\vec{x})$$
$$\left(-i(\gamma^0\partial_0 + \gamma^k\partial_k) + \frac{mc}{\hbar}\right)\psi(t,\vec{x}) = \left(-i\gamma^\mu\partial_\mu + \frac{mc}{\hbar}\right)\psi(t,\vec{x}) = 0 \tag{9.29}$$

We will show in Sec. 9.4 that this equation not only looks invariant due to its suggestive notation, but is indeed relativistically invariant, i.e., remains valid in any reference frame. Properties of the γ-matrices are obtained form those of α^k and β:

$$\{\gamma^\mu,\gamma^\nu\} = \gamma^\mu\gamma^\nu + \gamma^\nu\gamma^\mu = 2g^{\mu\nu}\mathbb{I} \tag{9.30}$$

and

$$\left(\gamma^0\right)^\dagger = \beta^\dagger = \gamma^0, \quad \left(\gamma^i\right)^\dagger = \left(\beta\alpha^i\right)^\dagger = \left(\alpha^i\right)^\dagger\beta^\dagger = \alpha^i\beta = -\gamma^i \tag{9.31}$$
$$\text{resp. } \left(\gamma^\mu\right)^\dagger = \gamma^0\gamma^\mu\gamma 0\,. \tag{9.32}$$

Their covariant counterpart is defined as for four-vectors of numbers $\gamma_\mu = g_{\mu\nu}\gamma^\nu$. Products of the from $\gamma^\mu x_\mu = \gamma^0 x^0 - \vec{\gamma}\vec{x}$ are often abbreviated to $\not{x}$. Note that

$$\not{\partial} = \gamma^\mu\partial_\mu = \frac{1}{c}\gamma^0\partial_t + \gamma^k\partial_k = \frac{1}{c}\gamma^0\frac{\partial}{\partial t} + \gamma^k\frac{\partial}{\partial x^k} \tag{9.33}$$

due to the "opposite" sign in the four gradient, see (9.4).

Finally, it is important to note that the individual γ matrices, as well as β and α do not act on four-vectors, but on "Dirac spinors". The fact that their dimension is four, like that of space-time, is a coincidence.

9.3.1.1 Electro-magnetic field

Coupling to a magneto-electric field is - as in the non-relativistic case - introduced via "minimal coupling", i.e.,

$$p_\mu \to p_\mu - \frac{e}{c}A_\mu, \quad \text{resp.} \quad i\hbar\partial_\mu \to i\hbar\partial_\mu - \frac{e}{c}A_\mu \tag{9.34}$$

where $A^\mu = (\Phi,\vec{A})$ combines scalar potential Φ and vector potential $\vec{A}$. This implies

$$i\hbar\partial_0 \to i\hbar\partial_0 - \frac{e}{c}\Phi, \quad \text{resp.} \quad i\hbar\partial_t \to i\hbar\partial_t - e\Phi \tag{9.35}$$
$$i\hbar\partial_k \to i\hbar\partial_k - \frac{e}{c}A_k, \quad \text{resp.} \quad -i\hbar\partial_k \to -i\hbar\partial_k - \frac{e}{c}A^k\,.$$

The covariant form of the Dirac equation becomes then

$$\left(-\gamma^\mu\left(i\hbar\partial_\mu - \frac{e}{c}A_\mu\right) + mc\right)\psi(t,\vec{x}) = 0 \tag{9.36}$$

9.3.2 Free particle and Non-relativistic limit

Only in few instances can the Dirac equation be solved, we will here discuss for illustration the free particle and the non-relativistic limit of a particle in an electro-magnetic field. The free particle will first be considered at rest, which identifies the rest mass as the dominant energy scale and introduces anti-particles. The non-relativistic limit can then be obtained in perturbation theory for energies small compared to the rest mass. Finally, the free particle with finite momentum is considered.

9.3.2.1 Free particle at rest – Negative-energy states

A free particle at rest sees no potential and has no kinetic energy, i.e., the "spatial" part $\vec{p}$ of the four-momentum vanishes. Consequently, $E^2 = m^2c^4$ according to (9.10) and the Dirac equation (9.20) becomes

$$i\hbar\frac{\partial}{\partial_t}\psi(t,\vec{x}) = \beta mc^2\psi(t,\vec{x}) = mc^2\begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&-1&0\\0&0&0&-1\end{pmatrix}\psi(t,\vec{x})\ . \tag{9.37}$$

As the operator on the right-hand side is already diagonal, there are clearly four solution with two energies, $E = \pm mc^2$:

$$u_1^+ = e^{-i\frac{mc^2}{\hbar}t}\begin{pmatrix}1\\0\\0\\0\end{pmatrix},\quad u_2^+ = e^{-i\frac{mc^2}{\hbar}t}\begin{pmatrix}0\\1\\0\\0\end{pmatrix},\quad u_1^- = e^{+i\frac{mc^2}{\hbar}t}\begin{pmatrix}0\\0\\1\\0\end{pmatrix},\quad u_2^- = e^{+i\frac{mc^2}{\hbar}t}\begin{pmatrix}0\\0\\0\\1\end{pmatrix}. \tag{9.38}$$

As in the case of the Klein-Gordon equation, there are solutions with negative energy. In the ground state, which is defined as the state with lowest energy, all these states are filled, because leaving one of them empty raises the energy. The important step in the interpretation of relativistic wave equations is to state that the vacuum corresponds to this ground state, i.e., is not empty but rather contains all negative-energy electrons. Observable particles are then on one hand positive-energy electrons and on the other hand empty negative-energy states. The latter also have positive energy. Particle creation can be understood as taking a negative-energy particle and moving it into an empty positive-energy state: In addition to the positive-energy particle, a (positive-energy) hole has been created, their total energy amounts to the energy difference between the two states. As the total energy of the pair is positive and twice that of the positive-energy particle, the hole can also be interpreted as a positive-energy "anti-particle".

9.3.2.2 Non-relativistic limit

When the speed of light is large compared to that of an electron, mc^2 is much larger than the kinetic energy. For this non-relativistic limit of small $|\vec{v}|/c$, perturbation theory starting from the four solutions $u^\pm_{1/2}$ seems consequently a good idea, where we focus here on the *high*-energy states u^+. We are interested in obtaining a low-energy Hamiltonian and start

from (9.20), where we identify βmc^2 as the exactly solved part H_0 and $V = -i\hbar c\alpha^k\partial_k$ as a perturbation.

For an electron coupled to an atom, e.g., the potential energy is of the same order of magnitude as the kinetic energy, i.e., likewise much smaller than the rest energy. Coupling (9.36) to an electro-magnetic field can then also be included into the perturbation, here written into the form of Eq. (9.20) by use of (9.35):

$$\begin{aligned}\left(i\hbar\frac{\partial}{\partial t} - (-|e|)\Phi\right)\psi(t\vec{x}) &= \left(c\big(-i\hbar\partial_k - \frac{e}{c}A^k\big)\alpha^k + \beta mc^2\right)\psi(t,\vec{x}) \\ i\hbar\frac{\partial}{\partial t}\psi(t\vec{x}) &= \underbrace{\Big(c\big(-i\hbar\partial_k - \frac{e}{c}A^k\big)\alpha^k + (-|e|)\Phi}_{=V} + \underbrace{\beta mc^2}_{=H_0}\Big)\psi(t,\vec{x})\ . \end{aligned} \tag{9.39}$$

It is now convenient to introduce two-spinors χ^+ and χ^- referring to the doublets with positive and negative energy, respectively. The matrices β and α^k are then also used in the representation (9.28) written in terms of 2×2 Pauli matrices and the (still exact) Dirac equation is:

$$i\hbar\frac{\partial}{\partial t}\begin{pmatrix}\chi^+\\ \chi^-\end{pmatrix} = \left(\begin{pmatrix}0 & c\vec{\pi}\cdot\vec{\sigma}\\ c\vec{\pi}\cdot\vec{\sigma} & 0\end{pmatrix} - |e|\Phi\mathbb{I} + mc^2\beta\right)\begin{pmatrix}\chi^+\\ \chi^-\end{pmatrix}\ , \tag{9.40}$$

where $\vec{\pi} = \vec{p} - \frac{e}{c}\vec{A}$.

In order to carry out the perturbation theory, projectors P and Q onto the positive- and negative-energy states are useful. In four-spinor and two-spinor notation, they are given by the 4×4 resp. 2×2 matrices

$$P = \begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&0&0\\0&0&0&0\end{pmatrix} = \begin{pmatrix}\mathbb{I}&0\\0&0\end{pmatrix} \quad\text{and}\quad Q = \begin{pmatrix}0&0&0&0\\0&0&0&0\\0&0&1&0\\0&0&0&1\end{pmatrix} = \begin{pmatrix}0&0\\0&\mathbb{I}\end{pmatrix}\ . \tag{9.41}$$

The 2×2 structure of the equation makes degenerate perturbation theory a bit simpler.

The first-order correction is then, according to (5.28),

$$\begin{aligned}\Delta^1 = PVP &= \begin{pmatrix}\mathbb{I}&0\\0&0\end{pmatrix}\left(\begin{pmatrix}0 & c\vec{\pi}\cdot\vec{\sigma}\\ c\vec{\pi}\cdot\vec{\sigma} & 0\end{pmatrix} - |e|\Phi\begin{pmatrix}\mathbb{I}&0\\0&\mathbb{I}\end{pmatrix}\right)\begin{pmatrix}\mathbb{I}&0\\0&0\end{pmatrix} = \\ &= \begin{pmatrix}\mathbb{I}&0\\0&0\end{pmatrix}\left(\begin{pmatrix}0 & 0\\ c\vec{\pi}\cdot\vec{\sigma} & 0\end{pmatrix} - |e|\Phi\begin{pmatrix}\mathbb{I}&0\\0&0\end{pmatrix}\right) = -|e|\Phi P\ . \end{aligned} \tag{9.42}$$

This is diagonal in spinor-space and does moreover not lift the degeneracy between the $E > 0$ states.

The first-order correction to the two-spinor χ^+ can most easily be obtained from (5.27) and is

$$\frac{Q}{E_n^0 - H_0}V\underbrace{P|n\rangle}_{=\chi^+} = \frac{1}{2mc^2}\begin{pmatrix}0&0\\0&\mathbb{I}\end{pmatrix}\begin{pmatrix}0 & c\vec{\pi}\cdot\vec{\sigma}\\ c\vec{\pi}\cdot\vec{\sigma} & 0\end{pmatrix}\begin{pmatrix}\chi^+\\0\end{pmatrix} = \begin{pmatrix}0\\ \frac{\vec{\pi}\cdot\vec{\sigma}}{2mc}\chi^+\end{pmatrix}\ . \tag{9.43}$$

Since $\vec{p} \approx m\vec{v}$ in the non-relativistic limit and $\vec{\pi}$ is typically of the same order as $\vec{p}$, this correction is smaller by a factor of $|\vec{v}|/c$ than the contribution χ^+ of the positive-energy states. It is consequently often called the "small component" as opposed to the "large component" χ^+.

Finally, the second-order energy correction (5.28) yields

$$\Delta^2 = P\frac{VQV}{E_n^0 - H_0}P = \frac{1}{2mc^2}\begin{pmatrix} c^2(\vec{\pi}\cdot\vec{\sigma})^2 & 0 \\ 0 & 0 \end{pmatrix} . \tag{9.44}$$

Taking zeroth to second order together, the effective Hamiltonian (5.28) acting on the two states with positive energy is

$$i\hbar\frac{\partial\chi^+}{\partial t} = H_{\text{eff}}\chi^+ = \left(mc^2 - |e|\Phi + \frac{(\vec{\pi}\cdot\vec{\sigma})^2}{2m}\right)\chi^+ . \tag{9.45}$$

The second-order contribution can be evaluated further using $(\vec{\sigma}\vec{a})(\vec{\sigma}\vec{b}) = \vec{a}\vec{b} + i\vec{\sigma}(\vec{a}\times\vec{b})$:

$$(\vec{\pi}\cdot\vec{\sigma})^2 = \vec{\pi}^2 + i\vec{\sigma}\cdot\left((\vec{p}+\frac{|e|}{c}\vec{A})\times(\vec{p}+\frac{|e|}{c}\vec{A})\right) = \vec{\pi}^2 + i\frac{|e|}{c}\vec{\sigma}\cdot\left(\vec{p}\times\vec{A} + \vec{A}\times\vec{p}\right) \tag{9.46}$$

In the cross products, we have to take into account that the differential operator $\vec{p} = -i\hbar\nabla$ acts on $\vec{A}$ as well as on the wave function to the right, if it stands to the left of $\vec{A}$. Applying the product rule gives $\nabla\times\vec{A} = \vec{B}$ for the term where it acts on $\vec{A}$ (and does not act on the wave function), while the term where it acts on the wave function (and not on $\vec{A}$) is for clarity switched around, giving it a minus sign:

$$i\frac{|e|}{c}\vec{\sigma}\cdot\left(\vec{p}\times\vec{A} + \vec{A}\times\vec{p}\right) = \frac{|e|\hbar}{c}\vec{\sigma}\cdot\left((\nabla\times\vec{A}) - \vec{A}\times\nabla + \vec{A}\times\nabla\right) = \frac{|e|\hbar}{c}\vec{\sigma}\cdot\vec{B} \tag{9.47}$$

The effective Hamiltonian has thus become the Pauli equation for a (non-relativistic) two-spinor

$$i\hbar\frac{\partial\chi^+}{\partial t} = \left(mc^2 - |e|\Phi + \frac{1}{2m}\left(\vec{p}+\frac{|e|}{c}\vec{A}\right)^2 + \frac{|e|\hbar}{2mc}\vec{\sigma}\cdot\vec{B}\right)\chi^+ . \tag{9.48}$$

Higher orders in perturbation theory lead to further corrections, e.g. spin-orbit coupling between spin and orbital angular momentum, which becomes relevant for heavy elements. This procedure is called "Foldy-Wouthuysen tranfsormation" and is most practically done along the lines of Sec. 5.2.2.4.

The result (9.48) strongly suggests that the Dirac equation indeed describes a particle with spin 1/2 like the electron. Moreover, the factor 1/2 that one would get by treating spin as a "half-integer orbital angular momentum", and which would be incorrect, is correctly absent, yielding the g-factor of the spin as $g = 2$. This can be seen more explicitly for the special choice $\vec{A} = \frac{1}{2}\vec{B}\times\vec{x}$ with constant $\vec{B}$, where the Pauli equation becomes

$$i\hbar\frac{\partial\chi^+}{\partial t} = \Big(mc^2 - |e|\Phi + \frac{\vec{p}^2}{2m}\frac{|e|^2}{2mc^2}\vec{A}^2 + \frac{|e|}{2mc}\big(\underbrace{\vec{p}\cdot\vec{A}}_{=\vec{A}\vec{p}-i\hbar(\nabla\vec{A})} + \vec{A}\cdot\vec{p}\big) + \frac{|e|\hbar}{2mc}\vec{\sigma}\cdot\vec{B}\Big)\chi^+ =$$

$$= \Big(mc^2 - |e|\Phi + \frac{\vec{p}^2}{2m}\frac{|e|^2}{2mc^2}\vec{A}^2 + \frac{|e|}{2mc}\underbrace{(\vec{B}\times\vec{x})\cdot\vec{p}}_{=(\vec{x}\times\vec{p})\cdot\vec{B}} + \frac{|e|\hbar}{2mc}\vec{\sigma}\cdot\vec{B}\Big)\chi^+ =$$

$$= \Big(mc^2 - |e|\Phi + \frac{\vec{p}^2}{2m}\frac{|e|^2}{2mc^2}\vec{A}^2 + \frac{|e|}{2mc}(\underbrace{\vec{x}\times\vec{p}}_{=\vec{L}} + \underbrace{\vec{\sigma}}_{=2\vec{S}})\cdot\vec{B}\Big)\chi^+ \,. \tag{9.49}$$

$\nabla\vec{A} = 0$ was used. The correct magnetic moment $(\vec{L} + 2\vec{S})$ arises for the electron, which differs from total angular momentum $\vec{J} = \vec{L} + \vec{S}$ obtained in Sec. 3.4.2.

9.3.2.3 Free particle with finite momentum

After obtaining the wave functions for a free particle at rest, one might obtain those of a moving free particle by going to a moving reference frame. We will here, however, obtain the wave function directly from the covariant form (9.29) of the Dirac equation.

In order to find the solutions, we make use of its known symmetries: It has to be translationally invariant, i.e., be an eigenstate of (2.15), it has to be invariant with respect to an analogous shift in time, and it has to be invariant under Lorentz transformations. These requirements are fulfilled by solutions of the form

$$\psi_k(t,\vec{x}) = \mathrm{e}^{-ikx}u_k = \mathrm{e}^{-i(k^0x^0 - \vec{k}\vec{x})}u_k \tag{9.50}$$

with $k^\mu = p^\mu/\hbar = (\frac{E}{\hbar c}, \vec{p}/\hbar)$, where all dependence on space and time is in the exponential and u_k does not depend on t or $\vec{x}$. (This is a similar argument and wave function as the one introduced in Bloch wave functions for periodic lattices, except that the latter are only eigenstates of translations by *lattice* vectors and time is not included.)

For $\vec{k} = 0$, solutions have to reduce to Eq. (9.38) for the free particle at rest, i.e., we expect four linearly independent solutions for each k, which we label $\psi^\pm_{r,k}(t,\vec{x})$ with $r = 1, 2$. All four, with positive or negative energy, can be written as in (9.50), however, the choice is often made to define all $k^0 = \frac{|E|}{c} > 0$ and make this up by a minus sign in the exponential of the $\psi^-_{r,k}$, i.e., by setting

$$\psi^-_{r,k} = e^{+ikx}u^-_{r,k} \,. \tag{9.51}$$

Alternatively, one might require instead $k_0 = -|E|$ for the negative-energy solutions. (The two approaches differ by a sign in the spatial phase, but as solutions for $\pm\vec{p}$ are degenerate anyway, this is not essential.)

Inserting (9.50) into the Dirac equation (9.29) yields

$$\Big(-i\gamma^\mu\partial_\mu + \frac{mc}{\hbar}\Big)\,\mathrm{e}^{\mp ik_\mu x^\mu}u^\pm_{r,k} = \Big(\mp\gamma^\mu k_\mu + \frac{mc}{\hbar}\Big)\,\mathrm{e}^{\mp ikx}u^\pm_{r,k} = 0 \tag{9.52}$$

from which we get the Dirac equation for $u^\pm_{r,k}$ as

$$\Big(\gamma^\mu k_\mu \mp \frac{mc}{\hbar}\Big)\,u^\pm_{r,k} = \Big(\not{k} \mp \frac{mc}{\hbar}\Big)\,u^\pm_{r,k} = 0 \,. \tag{9.53}$$

The Klein-Gordon equation, which the solution should likewise fulfill, reads

$$\Big(\partial_\mu\partial^\mu + (\frac{mc}{\hbar})^2\Big)\,\mathrm{e}^{\mp ikx}u^\pm_{r,k} = 0 \quad \Rightarrow \quad \Big(-k_\mu k^\mu + (\frac{mc}{\hbar})^2\Big)\,u^\pm_{r,k} = 0 \,. \tag{9.54}$$

Since $k_\mu k^\mu - \frac{m^2c^2}{\hbar^2}$ is a number and not an operator, this implies $k_\mu k^\mu - \frac{m^2c^2}{\hbar^2} = 0$. Moreover, it turns out that

$$\not{k}\not{k} = \sum_{\mu,\nu} k_\mu \gamma^\mu k_\nu \gamma^\nu = \frac{1}{2} \sum_{\mu,\nu} k_\mu k_\nu \underbrace{(\gamma^\mu \gamma^\nu + \gamma^\nu \gamma^\mu)}_{=2g^{\mu\nu}} = k_\mu k^\mu \tag{9.55}$$

and consequently

$$\left(\not{k} - \frac{mc}{\hbar}\right)\left(\not{k} + \frac{mc}{\hbar}\right) = \left(k_\mu k^\mu - \frac{m^2c^2}{\hbar^2}\right) = 0\,. \tag{9.56}$$

Any state obtained by applying $\not{k} + \frac{mc}{\hbar}$ ($\not{k} - \frac{mc}{\hbar}$) to some spinor consequently fulfills (9.53) for $E > 0$ ($E < 0$) and a solution that reverts to (9.38) for $\vec{k} = 0$ is

$$u^\pm_{r,k} = c_N \left(\not{k} \pm \frac{mc}{\hbar}\right) u^\pm_{r,\vec{k}=0}\,, \tag{9.57}$$

where c_N is a normalization.

Again, it is helpful to introduce two-spinors $\chi_1 = \binom{1}{0}$ and $\chi_2 = \binom{0}{1}$ to write

$$u^+_{r,\vec{k}=0} = \begin{pmatrix} \chi_r \\ 0 \end{pmatrix} \quad \text{and} \quad u^-_{r,\vec{k}=0} = \begin{pmatrix} 0 \\ \chi_r \end{pmatrix}\,. \tag{9.58}$$

This leads to

$$\begin{aligned} u^+_{r,k} &= c_N \left[k_0 \begin{pmatrix} \mathbb{I} & 0 \\ 0 & -\mathbb{I} \end{pmatrix} - k^i \begin{pmatrix} 0 & \sigma^i \\ -\sigma^i & 0 \end{pmatrix} + \frac{mc}{\hbar} \begin{pmatrix} \mathbb{I} & 0 \\ 0 & \mathbb{I} \end{pmatrix}\right] \begin{pmatrix} \chi_r \\ 0 \end{pmatrix} = \\ &= c_N \begin{pmatrix} \left(\frac{E}{\hbar c} + \frac{mc}{\hbar}\right) & \chi_r \\ \frac{\vec{p}\vec{\sigma}}{\hbar} & \chi_r \end{pmatrix} \end{aligned} \tag{9.59}$$

and

$$\begin{aligned} u^-_{r,k} &= c_N \left[\frac{-|E|}{c\hbar} \begin{pmatrix} \mathbb{I} & 0 \\ 0 & -\mathbb{I} \end{pmatrix} + k^i \begin{pmatrix} 0 & \sigma^i \\ -\sigma^i & 0 \end{pmatrix} + \frac{mc}{\hbar} \begin{pmatrix} \mathbb{I} & 0 \\ 0 & \mathbb{I} \end{pmatrix}\right] \begin{pmatrix} 0 \\ \chi_r \end{pmatrix} = \\ &= c_N \begin{pmatrix} \frac{\vec{p}\vec{\sigma}}{\hbar} & \chi_r \\ \left(\frac{|E|}{\hbar c} + \frac{mc}{\hbar}\right) & \chi_r \end{pmatrix}\,. \end{aligned} \tag{9.60}$$

It should be noted that each of these solutions is still twice degenerate, because states with $\vec{p}$ and $-\vec{p}$ have the same energy, replacing $\vec{p}$ by $-\vec{p}$ in the above equations consequently yields valid solutions as well. (If the ansatz e^{-ikx} is chosen for the u^-, this other solution is found.)

A plausible choice for the normalization would be to set $\psi(t,\vec{x})^\dagger \psi(t,\vec{x}) = 1$. This would result in

$$1 = \frac{c_N^2}{\hbar^2}\Big(\underbrace{(\vec{p}\vec{\sigma})^2}_{=\vec{p}^2 = \frac{E^2}{c^2} - m^2c^2} + \Big(\frac{|E|}{c} + mc\Big)^2\Big) = \frac{c_N^2}{\hbar^2} \frac{2|E|}{c} \left(\frac{|E|}{c} + mc\right)$$

and lead to

$$c_N = \frac{\hbar}{\sqrt{2\frac{|E|}{c}\left(\frac{|E|}{c}+mc\right)}} \,. \tag{9.61}$$

However, this is not the choice usually made, because the continuity equation (9.26) can be written in a relativistically invariant form, if the density $\psi(t,\vec{x})^\dagger\psi(t,\vec{x})$ transforms like the 0^{th} component of a four-vector, see Sec. 9.5.3, rather than being invariant. One can understand this by noting that the density is "(rest) mass per volume" and if the (total rest) mass of a particle is to remain the same, the density has to grow when the volume contracts in a Lorentz transformation.

This relativistic normalization can be achieved by introducing the adjoint spinor

$$\bar{\psi} = \psi^\dagger\gamma^0 \tag{9.62}$$

and requesting $\bar{\psi}(t,\vec{x})\psi(t,\vec{x}) = 1$. The adjoint spinor is here

$$\begin{aligned}\bar{u}^\pm_{r,k} &= \left(u^\pm_{r,k}\right)^\dagger\gamma^0 = c_N(u^\pm_{r,\vec{k}=0})^\dagger\Big(\underbrace{(\gamma^\mu)^\dagger}_{=\gamma^0\gamma^\mu\gamma^0}k_\mu \pm \frac{mc}{\hbar}\Big)\gamma^0 = c_N\underbrace{(u^\pm_{r,\vec{k}=0})^\dagger\gamma^0}_{=\bar{u}^\pm_{r,\vec{k}=0}}\Big(\gamma^\mu k_\mu \pm \frac{mc}{\hbar}\Big) = \\ &= c_N\bar{u}^\pm_{r,\vec{k}=0}\Big(\not{k} \pm \frac{mc}{\hbar}\Big)\end{aligned} \tag{9.63}$$

which leads to the normalization requirement

$$\begin{aligned}1 &= c_N^2\bar{u}^\pm_{r,\vec{k}=0}\Big(\not{k} \pm \frac{mc}{\hbar}\Big)^2 u^\pm_{r,\vec{k}=0} = c_N^2\bar{u}^\pm_{r,\vec{k}=0}\Big(2\frac{m^2c^2}{\hbar^2} \pm 2\frac{mc}{\hbar}\not{k}\Big)u^\pm_{r,\vec{k}=0} = \\ &= c_N^2 2\frac{mc}{\hbar^2}\bar{u}^\pm_{r,\vec{k}=0}\Big(mc \pm \gamma^0\hbar k_0 \mp \gamma^k\hbar k_k\Big)u^\pm_{r,\vec{k}=0} = c_N^2 2\frac{mc}{\hbar^2}\bar{u}^\pm_{r,\vec{k}=0}\Big(mc + \frac{|E|}{c}\Big)u^\pm_{r,\vec{k}=0} = \\ &= c_N^2 2\frac{mc\big(mc+\frac{|E|}{c}\big)}{\hbar^2} \,.\end{aligned} \tag{9.64}$$

Contributions from γ^k drop out, as these matrices do not have diagonal entries. The normalization is then in fact

$$c_N = \frac{\hbar}{\sqrt{2mc\big(mc+\frac{|E|}{c}\big)}} \,. \tag{9.65}$$

With this choice, the density becomes $\psi(t,\vec{x})^\dagger\psi(t,\vec{x}) = |E|/(mc^2)$ and as the energy is the 0^{th} component of the four-momentum, it turns out to indeed transform like a 0^{th} component.

Finally, it is worth noting that $\bar{u}^\pm_{1,k}u^\pm_{2,k} = 0$, as the $(1,2)$ element of $\gamma^\mu k_\mu$ vanishes, and

$$\bar{u}^+_{r,k}u^-_{s,k} \propto \bar{u}^+_{r,\vec{k}=0}\underbrace{\Big(\not{k}+\frac{mc}{\hbar}\Big)\Big(\not{k}-\frac{mc}{\hbar}\Big)}_{=k^2-\frac{m^2c^2}{\hbar^2}=0}u^-_{s,\vec{k}=0} = 0 \,. \tag{9.66}$$

Spinors $u^\pm_{r,k}$ with different k are not orthogonal, but the full plane-wave solutions are, due to the exponential prefactor; yielding for $\sigma,\sigma' = \pm$ the relation $\bar{\psi}^\sigma_{r,k}(t,\vec{x})\psi^{\sigma'}_{s,k'}(t,\vec{x}) = \delta_{r,s}\delta_{\sigma,\sigma'}\delta(k-k')$.

9.4 Lorentz invariance of the Dirac equation

When the Dirac equation for a particle in a field is transformed into a moving reference frame, the equations describing the fields will obviously also have to be transformed. So far, everything is just as it was in Chap. 2 for translational and rotational invariance. We had there seen that a translated (rotated) wave function is equivalent to the original wave function with an opposite translation (rotation) applied to its spatial coordinates, see Eqs. (2.8) and (2.22). It turns out that the situation for four-spinors is somewhat less straightforward: Transforming the space-time coordinates entering the wave function via Eq. (9.11) is not enough, an additional transformation has to be applied to the spinor.

Let us first analyze the additional transformation of spinors. It is supposed to give us the transformed wave function $\psi'(x') = \psi'(ct', \vec{x}')$ in a transformed system starting from $\psi(x) = \psi(ct, \vec{x})$ of the original frame. Since quantum mechanics is a linear theory, i.e., a linear combination of valid wave functions is also a valid wave function, the mapping has to be linear $\psi' = S \cdot \psi$. The connection should moreover work locally which implies that it must not depend on t and $\vec{x}$. The mapping is consequently multiplication of the spinor ψ with a matrix S that depends on Λ:

$$\psi'(x') = \psi'(\Lambda x) = S(\Lambda)\psi(x) = S(\Lambda)\psi(\Lambda^{-1}x') \tag{9.67}$$

Note that the matrix S acts in spinor-space and not in space-time as Λ does! Comparison of the first and last part with Eqs. (2.8) and (2.22) reveals the new aspect: S was $S = 1$ in Chap. 2. Indeed, nothing else would be possible for a one-component wave function and the added flexibility of $S \neq 1$ is what allows the vector-valued Dirac equation to be relativistically invariant.

Lorentz invariance of the Dirac equation (9.29) now requires that

$$\left(-i\gamma^\nu \partial'_\nu + \frac{mc}{\hbar}\right) \psi'(x') = 0 \; . \tag{9.68}$$

the γ matrices are constant matrices in spinor space and are not supposed to change with the transformation. The same form should be obtained if we start from the equation in the original unprimed frame and reformulate it in primed quantities:

$$\begin{aligned} \left(-i\gamma^\mu \partial_\mu + \frac{mc}{\hbar}\right) \psi(x) &= \left(-i\gamma^\mu \Lambda^\nu{}_\mu \partial'_\nu + \frac{mc}{\hbar}\right) S^{-1}\psi'(x') = 0 \\ \left(-iS\gamma^\mu S^{-1}\Lambda^\nu{}_\mu \partial'_\nu + \underbrace{SS^{-1}}_{=\mathbb{I}} \frac{mc}{\hbar}\right) \psi'(x') &= 0 \end{aligned} \tag{9.69}$$

This has the same form as (9.68) if

$$\begin{aligned} S\gamma^\mu S^{-1}\Lambda^\nu{}_\mu = \gamma^\nu \quad &\text{or} \\ \Lambda^\nu{}_\mu \gamma^\mu = S^{-1}\gamma^\nu S \; . \end{aligned} \tag{9.70}$$

Note that Λ commutes with S and γ^μ, because its acts on space-time, while S and γ^μ act on spinors. If such an $S = S(\Lambda)$ can be found, the Dirac equation is invariant.

Apart from fulfilling (9.70), another property of S has to be that it leaves scalar products resp. norms invariant, i.e. $\bar{\psi}'\psi' = \bar{\psi}\psi$:

$$\bar{\psi}'\psi' = \psi'^\dagger \gamma^0 \psi' = \psi^\dagger S^\dagger \gamma^0 S\psi = \bar{\psi}\psi = \psi^\dagger \gamma^0 \psi \quad \Rightarrow$$

$$S^\dagger \gamma^0 S = \gamma^0 \quad \Rightarrow \quad \det S^\dagger \cdot \det S = 1 \tag{9.71}$$

imply $|\det S| = 1$, the arbitrary phase can be transformed away. [1] This is the relation for S that takes the role equivalent to $U^\dagger U = \mathbb{I}$ in "usual" symmetries.

To obtain $S(\Lambda)$ from Λ, we use an approach we had used before:

- Consider infinitesimal transformations Λ
- Do an expansion of Λ and S up to first order in the "small transformation"
- Figure out the infinitesimal S for the infinitesimal Λ (see below)
- Integrate both Λ and S to obtain finite transformations in space-time and spinor space, resp.

This approach works for Lorentz transformations that can be connected smoothly to the identity, i.e., not for "big changes" like time inversion. We will here consider only such transformations, and moreover keep to $\det\Lambda = 1$ and $\Lambda^0{}_0 \geq 1$, i.e., neither space nor time are inverted.

Up to first order in $\Delta\omega$, the infinitesimal $\Lambda^\nu{}_\mu = \mathbb{I} + \Delta\omega^\nu{}_\mu = g^\nu{}_\mu + \Delta\omega^\nu{}_\mu$ has to fulfill (9.14):

$$\begin{aligned} \left(g^\rho{}_\mu + \Delta\omega^\rho{}_\mu\right) g^{\mu\nu} \left(g^\sigma{}_\nu + \Delta\omega^\sigma{}_\nu\right) &= g^{\rho\sigma} \\ g^\rho{}_\mu g^{\mu\nu} g^\sigma{}_\nu + g^\rho{}_\mu g^{\mu\nu} \Delta\omega^\sigma{}_\nu + \Delta\omega^\rho{}_\mu g^{\mu\nu} g^\sigma{}_\nu + \mathcal{O}(\Delta\omega^2) &= g^{\rho\sigma} + g^{\rho\nu}\Delta\omega^\sigma{}_\nu + \Delta\omega^\rho{}_\mu g^{\mu\sigma} = \\ &= g^{\rho\sigma} + \Delta\omega^{\sigma\rho} + \Delta\omega^{\rho\sigma} \,, \end{aligned} \tag{9.72}$$

which means that the infinitesimal part has to be antisymmetric with $\Delta\omega^{\sigma\rho} = -\Delta\omega^{\rho\sigma}$. This leaves six independent non-zero matrix elements for $\Delta\omega$, corresponding to three purely spatial rotations around the x-, y-, and z-axes in addition to three so-called "boosts" that transform to finite velocity along x, y and z.

We now assume that $S(\Lambda)$ can similarly be expanded in a Taylor series, which we stop after first order,

$$S = \mathbb{I} + \tau, \quad \Rightarrow \quad S^{-1} = \mathbb{I} - \tau \,, \tag{9.73}$$

and insert into (9.70).

$$\begin{aligned} (\mathbb{I} - \tau)\gamma^\nu(\mathbb{I} + \tau) &= g^\nu{}_\mu \gamma^\mu + \Delta\omega^\nu{}_\mu \gamma^\mu \\ \gamma^\nu - \tau\gamma^\nu + \gamma^\nu\tau + \mathcal{O}(\tau^2) &= \gamma^\nu + \Delta\omega^\nu{}_\mu \gamma^\mu \\ \gamma^\nu\tau - \tau\gamma^\nu &= \Delta\omega^\nu{}_\mu \gamma^\mu \end{aligned} \tag{9.74}$$

determines τ up to multiples of $\mathbb{I}$, because the difference between two solutions commutes with all γ^ν, which makes it $\propto \mathbb{I}$. Up to first order in τ, $\det S = 1$ leads to

$$\det(\mathbb{I} + \tau) = \det \mathbb{I} + \operatorname{tr} \tau \quad \Rightarrow \quad \operatorname{tr} \tau = 0 \,. \tag{9.75}$$

This fixes τ and helps in finding solutions because a lot is known about traceless 4×4 matrices in the context of the γ^μ, and yields

$$\tau = \frac{1}{8}\Delta\omega^{\mu\nu}\left(\gamma_\mu\gamma_\nu - \gamma_\nu\gamma_\mu\right) = \frac{-i}{4}\Delta\omega^{\mu\nu}\sigma_{\mu\nu} \tag{9.76}$$

[1] Requiring $\psi'^\dagger\psi' = \psi^\dagger\psi$ would here in fact lead to the same restriction on $\det S$.

with

$$\sigma_{\mu\nu} = \frac{i}{2}\left(\gamma_\mu\gamma_\nu - \gamma_\nu\gamma_\mu\right) \ . \tag{9.77}$$

We will now discuss two examples, a spatial rotation around z and a boost to finite velocity along x.

9.5 Behavior of a Spinor under Lorentz Transformations

9.5.1 Behavior of a Spinor under rotations

Let us assume that we rotate the coordinate system around the z-axis. The coordinates of a fixed point in space the change as

$$\begin{aligned} x^1 = x \quad &\rightarrow \quad x'^1 = x + \Delta\phi y = x^1 + \Delta\phi x^2 \\ x^2 = y \quad &\rightarrow \quad x'^2 = y - \Delta\phi x = x^2 - \Delta\phi x^1 \\ x^3 = z = z' = x'^3, \qquad & x^0 = ct = ct' = x'^0 \ . \end{aligned} \tag{9.78}$$

The changes under this "passive" transformation are opposite to those one gets for the "active" transformation of a point moving in a fixed coordinate system, as we had discussed in the context of translations in (2.8).

In the language of infinitesimal $\Delta\omega^{\mu\nu}$, only $\Delta\omega^1{}_2 = -\Delta\omega^{12} = \Delta\omega_{21} = \Delta\phi$ are non-zero. Equation (9.76) then gives

$$\begin{aligned} \tau^{\rm rot} &= -\frac{\Delta\phi}{8}\left(\gamma_1\gamma_2 - \gamma_2\gamma_1\right) + \frac{\Delta\phi}{8}\left(\gamma_2\gamma_1 - \gamma_1\gamma_2\right) = -\frac{\Delta\phi}{4}\left(\gamma_1\gamma_2 - \gamma_2\gamma_1\right) = \\ &= -\frac{\Delta\phi}{4}\left(\begin{pmatrix} 0 & \sigma^1 \\ -\sigma^1 & 0 \end{pmatrix}\begin{pmatrix} 0 & \sigma^2 \\ -\sigma^2 & 0 \end{pmatrix} - \begin{pmatrix} 0 & \sigma^2 \\ -\sigma^2 & 0 \end{pmatrix}\begin{pmatrix} 0 & \sigma^1 \\ -\sigma^1 & 0 \end{pmatrix}\right) = \\ &= -\frac{\Delta\phi}{4}\begin{pmatrix} -\sigma^1\sigma^2 + \sigma^2\sigma^1 & 0 \\ 0 & -\sigma^1\sigma^2 + \sigma^2\sigma^1 \end{pmatrix} = \frac{i\Delta\phi}{2}\begin{pmatrix} \sigma^3 & 0 \\ 0 & \sigma^3 \end{pmatrix} = \frac{i\Delta\phi}{2}\Sigma^3 \ . \end{aligned} \tag{9.79}$$

Both the positive- and the negative-energy spinor turn out to transform under rotations as a non-relativistic Pauli spinor would, where σ^3 is the generator of infinitesimal rotations around z, see Sec. 3.1.3. The two entries referring to positive energy thus indeed refer to the two degrees of freedom of a spin 1/2. The transformations for a finite angle ϕ are analogous to (3.18)

$$S(\phi) = \mathrm{e}^{\frac{i\Delta\phi}{2}\Sigma^3} = \cos\frac{\phi}{2} + i\sin\frac{\phi}{2}\Sigma^3 \ , \tag{9.80}$$

except for the opposite sign due to the choice of a passive transformation.

The next goal is to find the full effect of a rotation onto a four-spinor; we need to combine operator S with the impact of the rotation of the position coordinates. We have already discussed the latter impact for non-relativistic wave functions in Eq. (2.22) and will proceed an an analogous manner here. The combined overall change of the spinor, which connects the first to the last part of Eq. (9.67), is now denoted by a linear operator (i.e. a matrix) $A(\Lambda)$, i.e., (with simplified notation $x' \rightarrow x$):

$$\psi'(x) = S(\Lambda)\psi(\Lambda^{-1}x) = A(\Lambda)\psi(x) \ . \tag{9.81}$$

For infinitesimal transformations, we can Taylor expand the spinor in order to find $A(\Lambda)$:

$$\begin{aligned}\psi'(x) &= (\mathbb{I}+\tau)\,\psi\left((\mathbb{I}-\Delta\omega)x\right) = (\mathbb{I}+\tau)\,\psi(x-\Delta\omega x) = (\mathbb{I}+\tau)\left(\psi(x)-(\Delta\omega x)^\nu\partial_\nu\psi(x)\right) = \\ &= \left(\mathbb{I}+\tau-(\Delta\omega x)^\nu\partial_\nu\right)\psi(x)+\mathcal{O}(\Delta\omega^2) = \left(\mathbb{I}+\tau-\Delta\omega^\nu{}_\mu x^\mu\partial_\nu\right)\psi(x)\,. \end{aligned} \tag{9.82}$$

Using (9.76) to express τ, one then finds

$$\psi'(x) = \left(\mathbb{I}-\frac{i}{4}\Delta\omega^{\mu\nu}\sigma_{\mu\nu}-\Delta\omega^{\nu\mu}x_\mu\partial_\nu\right)\psi(x) = \left(\mathbb{I}+\Delta\omega^{\mu\nu}\left(\frac{-i}{4}\sigma_{\mu\nu}+x_\mu\partial_\nu\right)\right)\psi(x) \tag{9.83}$$

for the full transformation from one reference system to another.

For the special case of a rotation around z, we have already evaluated the first part (coming from τ) in Eq. (9.79); the second part becomes

$$\begin{aligned}-\Delta\omega^{\mu\nu}x_\mu\partial_\nu &= -\Delta\omega^{1,2}x_1\partial_2+\Delta\omega^{2,1}x_2\partial_1 = -\Delta\phi\left(-x\partial_y+y\partial_x\right) = i\Delta\phi\big(x(-i\partial_y)-y(-i\partial_x)\big) = \\ &= i\Delta\phi L^3 \end{aligned} \tag{9.84}$$

yielding a total

$$\mathbb{I}+\tau+\Delta\omega^{\mu\nu}x_\mu\partial_\nu = \mathbb{I}+i\Delta\phi\left(\frac{1}{2}\Sigma^3+L^3\right) = \mathbb{I}+i\Delta\phi\;J^3\,. \tag{9.85}$$

Analogous results are of course obtained for rotations around the x and y axes: Rotation of a spinful electron is thus generated by the total angular momentum $\vec{J}=\vec{L}+\vec{S}$. Orientation and rotation in spins space are no longer purely abstract but are here tied to rotations of the spinor around spatial directions.

9.5.2 Behavior of a Spinor under a Boost

Transformation to a reference frame moving with infinitesimal velocity $-dv$ along x, called a boost, gives

$$\begin{aligned} x=x^1 \quad &\rightarrow \quad x'^1 = x-dvt = x^1-\frac{dv}{c}x^0 \\ &\Rightarrow \quad \Delta\omega^1{}_0 = \Delta\omega^{10} = -\frac{dv}{c} = -\Delta\beta\,. \end{aligned} \tag{9.86}$$

For the transformation τ, one finds

$$\tau^{\text{boost}} = -\frac{\Delta\beta}{8}\left(\gamma_1\gamma_0-\gamma_0\gamma_1\right)+\frac{\Delta\beta}{8}\left(\gamma_0\gamma_1-\gamma_1\gamma_0\right) = \frac{\Delta\beta}{2}\alpha_1\,. \tag{9.87}$$

For a particle moving along x, the first and the last spinor component are thus connected, this agrees with the spinor of the moving free Dirac particle, where p^x enters in the same place, see (9.59).

9.5.3 Adjoint Spinor and Covariant Continuity Equation

The adjoint spinor was introduced above, and the normalization chosen accordingly, to make the density proportional to the energy rather than constant. The same choice also makes the components of the current into components of a three-vector that can be combined with the density to a four vector. The continuity equation (9.26) is here rewritten in terms of γ-matrices:

$$\begin{aligned}
&\frac{\partial}{\partial t}\left(\psi^\dagger(t,\vec{x})\gamma^0\gamma^0\psi(t,\vec{x})\right) = -c\partial_k\left(\psi^\dagger(t,\vec{x})\gamma^0\gamma^0\alpha^k\psi(t,\vec{x})\right)\\
&\frac{\partial}{\partial t}\left(\bar{\psi}(t,\vec{x})\gamma^0\psi(t,\vec{x})\right) + c\partial_k\left(\bar{\psi}(t,\vec{x})\gamma^k\psi(t,\vec{x})\right) =\\
&= \partial_\mu\left(c\bar{\psi}(t,\vec{x})\gamma^\mu\psi(t,\vec{x})\right) = \partial_\mu j^\mu = 0\,. \qquad (9.88)
\end{aligned}$$

This is certainly a nicely compact form of the continuity equation, but we still need to verify that j^μ transforms as a four vector.

Under a Lorentz transformation together with a transformation S of the spinor, the current becomes

$$\begin{aligned}
j'^\mu(x')/c &= \bar{\psi}'(x')\gamma^\mu\psi'(x') = \psi'^\dagger(x')\gamma^0\gamma^\mu\psi'(x') = \psi(x)^\dagger S^\dagger\gamma^0\gamma^\mu S\psi(x) =\\
&= \psi(x)^\dagger\gamma^0 S^{-1}\gamma^\mu S\psi(x) = \bar{\psi}(x)S^{-1}\gamma^\mu S\psi(x) = \bar{\psi}(x)\Lambda^\mu{}_\nu\gamma^\nu\psi(x) = \Lambda^\mu{}_\nu\bar{\psi}(x)\gamma^\nu\psi(x) =\\
&= \Lambda^\mu{}_\nu j^\nu(x)/c\,. \qquad (9.89)
\end{aligned}$$

It turns out to indeed transform as in (9.11), i.e., like a four-vector. To find this result, we have used $S^\dagger\gamma^0 = \gamma^0 S^{-1}$, which in turn comes out of requiring that the norm defined as $\bar{\psi}'\psi' = \bar{\psi}\psi$ be unchanged under Lorentz transformations, see (9.71).

9.5.4 Symmetries: Angular momentum, Helicity and chirality

9.5.4.1 Spin and helicity

Without an electromagnetic field, the positive- and negative-energy doublets of a free Dirac-fermion were degenerate, even at at finite momentum. Based on the non-relativistic limit discussed in Sec. 9.3.2.2 and on the behaviour under rotation discussed in Sec. 9.5.1, the two states of the doublet can be concluded to refer to two spin projections. However, spin orientation is tied to directions in space for relativistic particle, see Eq. (9.85), and this has consequences for symmetries: While space looks the same in all directions for a particle at rest, a moving particle has a special direction, namely that of its momentum.

This reduced symmetry suggests that only rotations around this special direction have to leave the system invariant, implying that spin only projection only has to be a good quantum number along this direction. We are going to show explicitly that this is indeed the case.

Focussing on the positive-energy solutions (everything is the same for the negative-energy ones), we first not that any linear combination of the two degenerate states is also a valid eigenstate. Looking at the spinors given in Eq. (9.59), one sees that such a linear combination amounts to choosing the two-spinor χ_r differently from that basis choices $\chi_1 = \binom{1}{0}$ and $\chi_2 = \binom{0}{1}$, i.e., any two-spinor can be used to define a solution to the free Dirac equation via inserting it into Eq. (9.59).

One 'obvious' choice for χ would be an eigenstate of $\vec{p}\vec{\sigma}$, resp. $\vec{k}\vec{\sigma}$. As we are here dealing with (Pauli-) two-spinors, we can use everything learned there: the two eigenstates would correspond to projections along $+\hat{k}$ and $-\hat{k}$, with $\hat{k} = \vec{k}/|k|$ being the normalized quantization axis. labelling such state as $|\sigma = \pm, \vec{k}\rangle$, Eq. (9.59) together with (9.65)becomes

$$u^{+}_{\sigma,\vec{k}} = \frac{1}{\sqrt{2mc}} \begin{pmatrix} \sqrt{\frac{E}{c} + mc} & |\sigma, \vec{k}\rangle \\ \frac{\hbar \vec{k}\vec{\sigma}}{\sqrt{\frac{|E|}{c} + mc}} & |\sigma, \vec{k}\rangle \end{pmatrix} . \tag{9.90}$$

Making use of the eigenvalue $\vec{k}\vec{\sigma}|\sigma = \pm, \vec{k}\rangle = \pm|k||\sigma, \vec{k}\rangle$ and of the energy-momentum relation $\frac{E^2}{c^2} - m^2c^2 = \hbar^2|k|^2$, this can be simplified to

$$u^{+}_{\sigma,\vec{k}} = \frac{1}{\sqrt{2mc}} \begin{pmatrix} \sqrt{\frac{E}{c} + mc} & |\sigma, \vec{k}\rangle \\ \pm\sqrt{\frac{E}{c} - mc} & |\sigma, \vec{k}\rangle \end{pmatrix} . \tag{9.91}$$

The spin projection along the propagation direction is thus conserved, the corresponding property is called 'helicity'. One sees that for $|k| \to 0$, and $E \to mc^2$, the only the upper two entries of the spinor $u^{+}_{\sigma,\vec{k}}$ remain. As $\vec{k}$ does then not enter $u^{+}_{\sigma,\vec{k}}$, we then regain the higher non-relativistic symmetry, where spin along any direction is conserved.

9.5.4.2 Total angular momentum

While tying the spin to a spatial direction removes a large part of the non-relativistic symmetry, so that only helicity survives, space for a free Dirac fermion is still isotropic. Accordingly, a 'complete' rotation of the particle, addressing both spin and position-space coordinates, should leave its description invariant. As we have seen in Sec. 9.5.1, the generator of such a rotation is total angular momentum, suggesting that it should be the conserved quantity in question.

Using the Hamiltonian

$$i\hbar\frac{\partial}{\partial t}\psi = \left(-i\hbar c\alpha^k\partial_k + \beta mc^2\right)\psi = \hat{H}_D\psi \, , \tag{9.92}$$

defined by the free Dirac equation (9.20), we can check whether its commutator with a given quantity vanishes. If so, a common eigenstate of both operators will remain one for all times, so that the other quantity is conserved, just as in the non-relativistic case.

Building on the calculations done for Eq. (9.79), one can express 4×4 matrices for the spin in terms of γ commutators:

$$\Sigma^3 = \begin{pmatrix} \sigma^3 & 0 \\ 0 & \sigma^3 \end{pmatrix} = i\gamma_1\gamma_2 \quad \text{and} \quad \Sigma^i = \begin{pmatrix} \sigma^i & 0 \\ 0 & \sigma^i \end{pmatrix} = \frac{i}{2}\epsilon^{ijk}\gamma_j\gamma_k \, . \tag{9.93}$$

That these matrices describe spin is clear in standard representation, where the Dirac four-spinor is composed of two Pauli two-spinors. The r.h.s. expressions in terms of γ matrices remain valid for any representation.

The commutator is then, using $\Sigma_i = -\Sigma^i$, found as

$$\left[\hat{H}_D, \Sigma_i\right] = -i\hbar c\left[\alpha^l, \Sigma_i\right]\partial_l + \left[\beta^0, \Sigma_i\right]mc^2 = -i\hbar c\left[\gamma^0\gamma^l, \Sigma_i\right]\partial_l + \left[\gamma^0, \Sigma_i\right]mc^2 =$$

$$
\begin{aligned}
&= -\frac{\hbar c}{2}\epsilon_{ijk}\left[\gamma^0\gamma^l, \gamma^j\gamma^k\right]\partial_l - \frac{i}{2}\epsilon_{ijk}\left[\gamma^0, \gamma^j\gamma^k\right]mc^2 = \\
&= -\frac{\hbar c}{2}\epsilon_{ijk}\left(\left[\gamma^0\gamma^l, \gamma^j\right]\gamma^k + \gamma^j\left[\gamma^0\gamma^l, \gamma^k\right]\right)\partial_l - \frac{i}{2}\epsilon_{ijk}\Big(\underbrace{\{\gamma^0, \gamma^j\}}_{=0}\gamma^k - \gamma^j\underbrace{\{\gamma^0, \gamma^k\}}_{=0}\Big)mc^2 = \\
&= -\frac{\hbar c}{2}\epsilon_{ijk}\Big(\gamma^0\underbrace{\{\gamma^l, \gamma^j\}}_{=2g^{l,j}}\gamma^k - \underbrace{\{\gamma^0, \gamma^j\}}_{=0}\gamma^l\gamma^k + \gamma^j\gamma^0\underbrace{\{\gamma^l, \gamma^k\}}_{=2g^{l,k}} - \gamma^j\underbrace{\{\gamma^0, \gamma^k\}}_{=0}\gamma^l\Big)\partial_l = \\
&= -\hbar c\epsilon_{ijk}\Big(g^{l,j}\gamma^0\gamma^k + \gamma^j\gamma^0 g^{l,k}\Big)\partial_l = -\hbar c\epsilon_{ijk}\Big(\gamma^0\gamma^k\partial^j + \gamma^j\gamma^0\partial^k\Big) = \\
&= 2c\hbar\epsilon_{ijk}\gamma^0\gamma^j\partial^k = 2ic\epsilon_{ijk}\gamma^0\gamma^j(-i\hbar\partial^k) = 2ic\epsilon_{ijk}\gamma^0\gamma^j p^k\ , \qquad (9.94)
\end{aligned}
$$

i.e., spin is indeed not conserved unless $\vec{p} = 0$.

The commutator with orbital angular momentum $L_i = \epsilon_{ijk}x^j p^k$ is then

$$
\begin{aligned}
&\left[\hat{H}_D, L_i\right] = -i\hbar c\epsilon_{ijk}\gamma_0\gamma_l\left[\partial^l, x^j p^k\right] + 0 = -c\epsilon_{ijk}\gamma_0\gamma_l\left[p^l, x^j p^k\right] = -c\epsilon_{ijk}\gamma_0\gamma_l\underbrace{\left[p^l, x^j\right]}_{=-i\hbar\delta_{l,j}=i\hbar g^{lj}}p^k \\
&= -i\hbar c\epsilon_{ijk}\gamma^0\gamma^j p^k \qquad (9.95)
\end{aligned}
$$

and likewise not a conserved quantity for $\vec{p} \neq 0$. However, the commutators for the combined quantity $L_i + \frac{\hbar}{2}\Sigma_i$ exactly cancel, so that total angular momentum is conserved. Note that for the particle at rest, $\vec{L}$ and $\vec{\Sigma}$ are conserved individualy.

9.5.4.3 Chirality

For the extremely relativistic limit, namely $mc \ll \frac{|E|}{c}$, yet a different symmetry emerges, called 'chirality'. It is expressed by the matrix

$$
\gamma^5 = \gamma_5 = \gamma_5^\dagger = i\gamma^0\gamma^1\gamma^2\gamma^3, \quad \text{with} \quad \gamma^5\gamma^\mu + \gamma^\mu\gamma^5 = \{\gamma^5, \gamma^\mu\} = 0 \quad \text{and} \quad \left(\gamma^5\right)^2 = \mathbb{I}\,, \tag{9.96}
$$

and is in standard representation given by

$$
\gamma^5 = \begin{pmatrix} 0 & \mathbb{I} \\ \mathbb{I} & 0 \end{pmatrix}\ . \tag{9.97}
$$

Its commutator with the Dirac Hamiltonian (9.92) is

$$
\begin{aligned}
\left[\gamma^5, \hat{H}_D\right] &= -i\hbar c\left[\gamma^5, \gamma^0\gamma^k\right]\partial_k + mc^2\left[\gamma^5, \gamma^0\right] = \\
&= -i\hbar c\Big(\{\gamma^5, \gamma^0\}\gamma^k - \gamma^0\{\gamma^5, \gamma^k\}\Big)\partial_k + 2mc^2\gamma^5\gamma^0 = 2mc^2\gamma^5\gamma^0\ . \qquad (9.98)
\end{aligned}
$$

The quantity expressed by γ^5 is thus conserved for $m = 0$. This limit is reached at very high energies, for fermions with (nearly) vanishing mass (neutrinos) or when the Dirac equation is used as an effective description of electronic states in graphene. One sees from Eq. (9.97) that eigenstates of γ^5 combine particle and anti-particle character: Once $m \approx 0$, particle and anti-particle are not longer clearly different so that a new symmetry can emerge. Note

that for $E \gg m$, Eq. (9.91) yields these chirality eigenstates, helicity and chirality can then be diagonalized together.

While we have so far used the "standard representation" of the γ matrices, one can perform a basis change $\tilde{\psi} = U^\dagger \psi$ in spinor-space, which changes the representation. An alternative choice is the "chiral" representation given by $U = \frac{1}{\sqrt{2}}(1 + \gamma^5\gamma^0)$. In this bases, the γ matrices become

$$\tilde{\gamma}^0 = U^\dagger \gamma^0 U = \frac{1}{2}(1 + \underbrace{(\gamma^5\gamma^0)^\dagger}_{=\gamma^0\gamma^5})\gamma^0(1+\gamma^5\gamma^0) = \frac{1}{2}(\gamma^0 + \gamma^0\gamma^5\gamma^0 + \gamma^0\gamma^5\gamma^0 + \underbrace{\gamma^0\gamma^5\gamma^0\gamma^5\gamma^0}_{=-\gamma^0}) = -\gamma^5 \tag{9.99}$$

$$\tilde{\gamma}^5 = \frac{1}{2}(1+\gamma^0\gamma^5)\gamma^5(1+\gamma^5\gamma^0) = \frac{1}{2}(\gamma^5 + \gamma^5\gamma^5\gamma^0 + \gamma^0\gamma^5\gamma^5 + \underbrace{\gamma^0\gamma^5\gamma^5\gamma^5\gamma^0}_{=-\gamma^5}) = \gamma^0 \tag{9.100}$$

$$\tilde{\gamma}^k = \frac{1}{2}(1+\gamma^0\gamma^5)\gamma^k(1+\gamma^5\gamma^0) = \frac{1}{2}(\gamma^k + \underbrace{\gamma^k\gamma^5\gamma^0 + \gamma^0\gamma^5\gamma^k}_{=\{\gamma^5,\gamma^0\}\gamma^k=0} + \underbrace{\gamma^0\gamma^5\gamma^k\gamma^5\gamma^0}_{=\gamma^k}) = \gamma^k \,. \tag{9.101}$$

The chirality is here diagonal.

Similar to Eq. (9.40), we can write the Dirac equation again in terms of two-spinors, which refer here to states with positive and negative chirality rather than positive and negative energy. Using

$$\tilde{\alpha}^k = \tilde{\gamma}^0\tilde{\gamma}^k = -\gamma^5\gamma^k = \begin{pmatrix} \sigma^k & 0 \\ 0 & -\sigma^k \end{pmatrix} \quad \text{and} \quad \tilde{\beta} = \tilde{\gamma}^0 = -\gamma^5 \tag{9.102}$$

the Dirac equation for a free fermion becomes in chiral representation

$$i\hbar\frac{\partial}{\partial t}\begin{pmatrix} \tilde{\chi}^+ \\ \tilde{\chi}^- \end{pmatrix} = \left(\begin{pmatrix} c\vec{p}\cdot\vec{\sigma} & 0 \\ 0 & -c\vec{p}\cdot\vec{\sigma} \end{pmatrix} - mc^2\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\right)\begin{pmatrix} \tilde{\chi}^+ \\ \tilde{\chi}^- \end{pmatrix} . \tag{9.103}$$

These new two-spinors $\tilde{\chi}^\pm$ with fixed chirality are called "Weyl spinors".

Compared to Eq. (9.40) in terms of Pauli spinors, the roles of $\vec{p}\vec{\sigma}$ and m are interchanged: For Pauli spinors, momentum mixes the states with positive and negative mass, i.e., Pauli spinors are only good approximate eigenstates if this mixing is small compared to the mass. For Weyl spinors, on the other hand, the kinetic energy preserves chirality and they are good approximate eigenstates if the mass, which mixes chiralities, is small. For neutrinos, "oscillations", where one kind of neutrino changes into another, work in a similar manner and indicate that the mass causing this mixing is small but not zero. Apart from neutrino physics, a Hamiltonian like (9.103) with $m = 0$ is currently also discussed in "Weyl semi metals", where it describes the electronic states near the Fermi level.